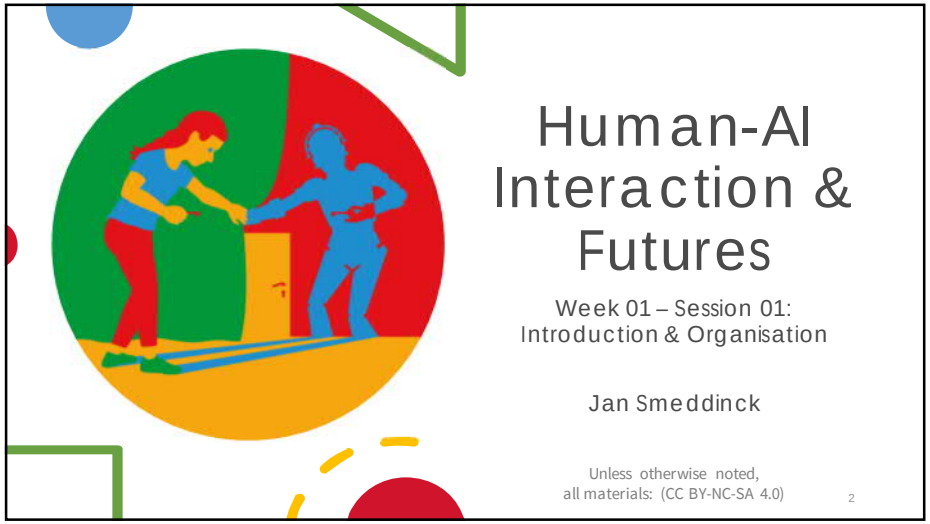
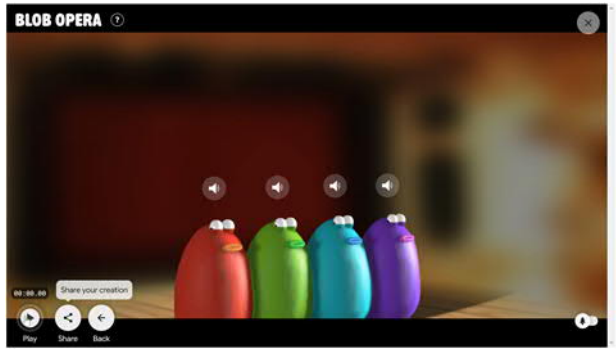
1



2

Blob Opera



https://g.co/arts/SxNT6ZsNAQrCgbjR6

3

3

Module Leaders

Jan Smeddinck
jan.smeddinck@ncl.ac.uk
https://openlab.ncl.ac.uk/people/jan-smeddick/

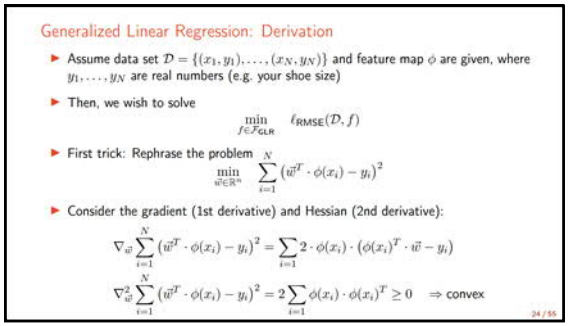
Yu Guan
yu.guan@ncl.ac.uk
https://openlab.ncl.ac.uk/people/yu-guan/

4

4

Motivation for this module…

- Benjamin Paaßen (Uni Sydney): Introduction to Machine Learning: Session 1: The Basics of Optimization, Probability, and Machine Learning



5

5

Motivation for this module…

And incredible developments are happening all the time (example of the day: GPT-3): https://www.youtube.com/watch?v=_x9AwxfjxvE

It takes about 10 minutes to make this:

6

6

Motivation for this module…

- Most AI/ML courses are disconnected from real-world problems…
- Most AI/ML courses consider "user-interfaces" or human impact as an afterthought; and focus narrowly on algorithms…
- *… why is this a problem?*

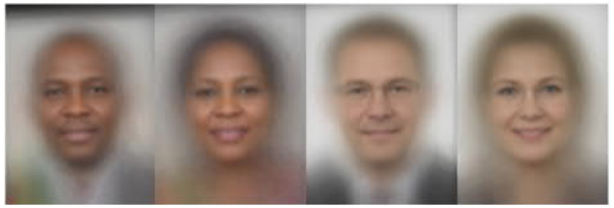
7

7

Gender Shades

Buolamwini, J., & Gebru, T. (2018). *Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification*. 15.

Gender Classifier	Darker Male	Darker Female	Lighter Male	Lighter Female	Largest Gap
Microsoft	94.0%	79.2%	100%	98.3%	20.8%
FACE++	99.3%	65.5%	99.2%	94.0%	33.8%
IBM	88.0%	65.3%	99.7%	92.9%	34.4%

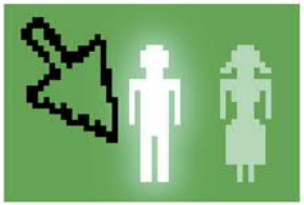


8

8

Algorithmic Bias

African-American defendants are substantially more likely than **White-American** defendants to be incorrectly classified as high risk by the algorithms.



Google Ads displayed more high paying jobs for **male users** than for **female users**.

https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

see: http://civilrightsdocs.info/pdf/criminal-justice/Pretrial-Risk-Assessment-Full.pdf ... via https://haiicmu.github.io

9

9

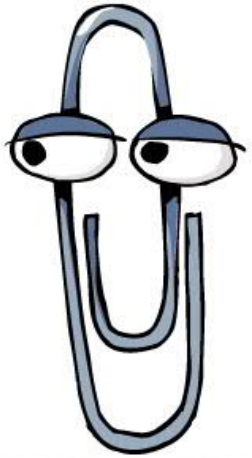
Lethal Autonomous Weapons (LAW)

- AI killer robots ... far fetched?
- E.g. Peter Asaro / Mark Gubrud: *any weapon system that is capable of releasing a lethal force without the operation, decision, or confirmation of a human supervisor can be deemed autonomous*
  - Asaro, Peter (2012). "On Banning Autonomous Weapon Systems: Human Rights, Automation, and the Dehumanization of Lethal Decision-Making". *Red Cross*. 687: 94.
  - "Autonomy without Mystery: Where do you draw the line?". *1.0 Human*. 2014-05-09. Retrieved 2018-06-08.
- Charta against autonomous weapons: https://futureoflife.org/open-letter-autonomous-weapons/

CC-BY-SA: Srđan Popović

10

10



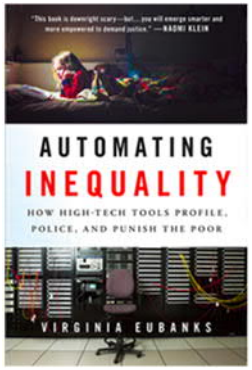
But it gets worse...

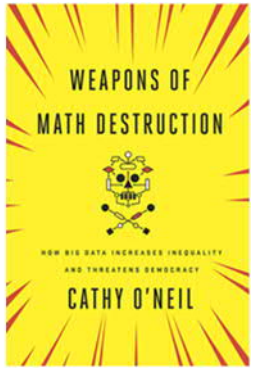
https://en.wikipedia.org/wiki/Office_Assistant

11

11

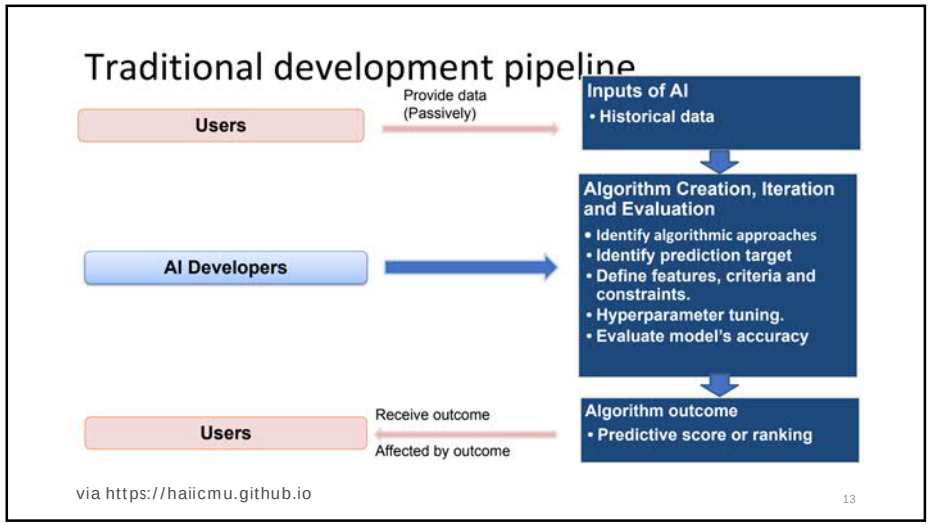
Further reading...



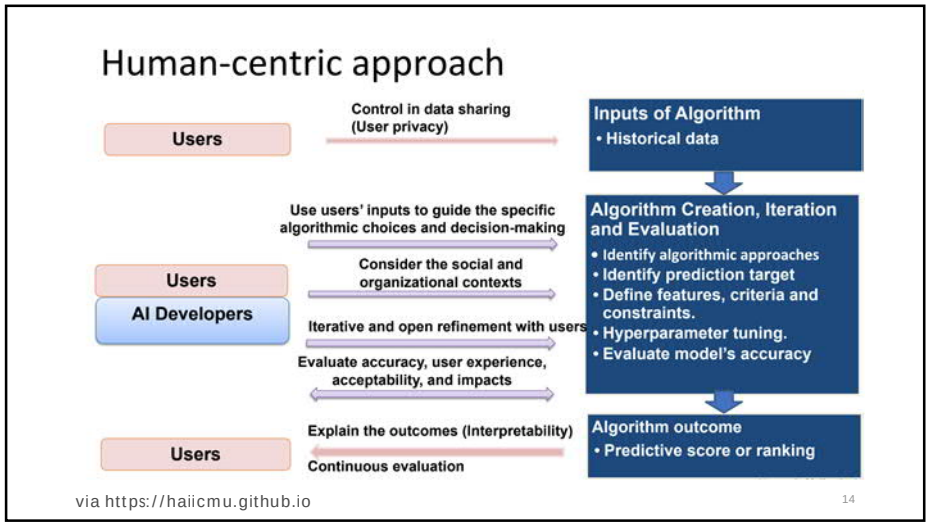


12

12



13



14

Acknowledgements

- This module builds on many existing sources and benefits from the generous contributions by a number of individuals…
- Key existing module: https://haiicmu.github.io by Haiyi Zhu and Steven Wu; formerly by Chinmay Kulkarni and Mary Beth Kery
- Individual contributions (lecture materials or sessions) by Alex Bowyer, Robert Porzel, Viana (Nijia) Zhang, and more.

15

15

Interdisciplinary College

Neuroscience · Neuroinformatics · Cognitive Science · Artificial Intelligence · and more

Annual interdisciplinary spring school …
… virtual (affordable!) version in 2021 (staring March 12).
Theme: *Connected in Cyberspace*

https://interdisciplinary-college.org/

16

16

Course Outline

- Week 1: Foundations of Human-AI Interaction
- Week 2: Operationalising AI in Practice
- Week 3: Advanced Topics and Futures of Human-AI Interaction
- Additional materials pages (Canvas), asynch comms (MS Teams), synch comms (MS Teams)

17

17

Course Outline: Week 01

- Topic 01 - Introduction & Organisation
- Topic 02 - Basic Concepts of AI
- Topic 03 - History of AI
- Topic 04 - Human-Data Interaction
- Topic 05 - Data Visualisation & Communication
- Topic 06 - Human-AI Interaction Design

18

18

Course Outline: Week 02

Yu Guan:

- Foundations of applied ML/AI …
- Examples around sensor (wearable) analytics (activity recognition) …
- Focus on supporting you with foundations for coursework

19

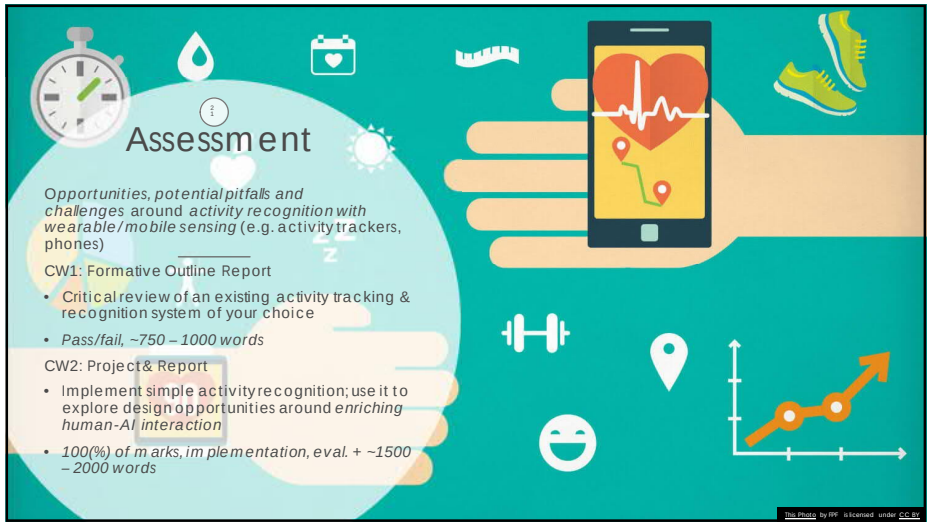
19

Course Outline: Week 03

- Topic 01 - Explainable, Interpretable & Relatable AI
- Topic 02 - AI Ethics
- Topic 03 - Humans-in-the-Loop
- Topic 04 - Recommender Systems
- Topic 05 - Conversational Interfaces
- Topic 06 - AI Agents & Robots
- Topic 07 - Human-AI Integration
- Topic 08 - Creative AI
- Topic 09 - Summary & Outlook

20

20



21

Deep Fake Sandbox

Fun video introduction link: https://www.youtube.com/watch?v=mUfJOQKdtAk

Siarohin, A., Lathuilière, S., Tulyakov, S., Ricci, E., & Sebe, N. (2019). First Order Motion Model for Image Animation. *Advances in Neural Information Processing Systems*, *32*, 7137–7147.

https://colab.research.google.com/github/AliaksandrSiarohin/first-order-model/blob/master/demo.ipynb

https://aliaksandrsiarohin.github.io/first-order-model-website/

22

22

Human-AI Interaction & Futures

Week 01 – Session 02:
Basic Concepts (ABC of AI)

Jan Smeddinck

Unless otherwise noted, all materials: (CC BY-NC-SA 4.0)

23

23

Learning Goals

Learn / get a refresher on the foundational terms behind…

- Artificial Intelligence
- Machine Learning
- (Non-Learning) Adaptive Systems
- … and their most general components / elements

24

24

This Photo by TODD JAQUITH is licensed under CC BY-NC

25

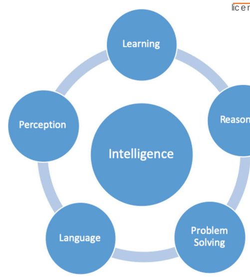
Artificial Intelligence

"… the designing and building of intelligent agents that receive percepts from the environment and take actions that affect that environment."

Russell, S., & Norvig, P. (2016). *Artificial Intelligence: A Modern Approach, Global Edition* (3rd edition). Pearson.

"Intelligence: The ability to learn and solve problems."

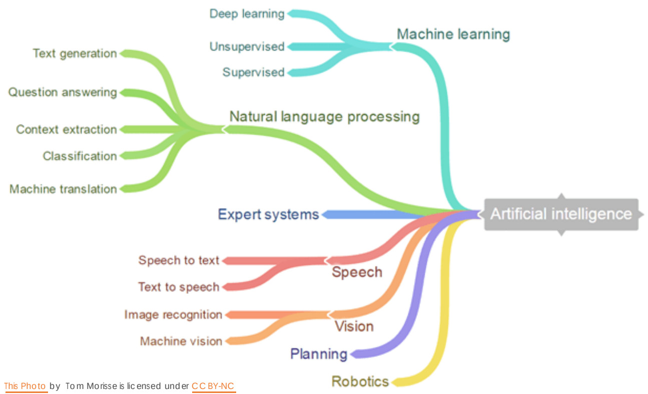
This Photo by Elliott Brown is licensed under CC BY



This Photo via Sessa, M., Khan, A. R., Liang, D., Andersen, M., & Kulahci, M. (2020). Artificial Intelligence in Pharmacoepidemiology: A Systematic Review. Part 1—Overview of Knowledge Discovery Techniques in Artificial Intelligence. Frontiers in Pharmacology, 11. is licensed under CC BY

26

Artificial Intelligence



This Photo by Tom Morisse is licensed under CC BY-NC

27

Machine Learning

A "machine" that is able to improve based on past experience without explicit human programming on how to improve each time.

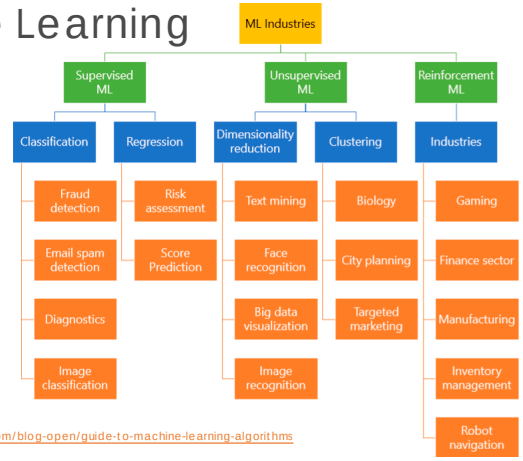
"A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P if its performance at tasks in T, as measured by P, improves with experience E."

-- Tom M. Mitchell

This Photo by World Economic Forum is licensed under CC BY-SA-NC

28

Machine Learning



Zee Gimon Via: https://huspi.com/blog-open/guide-to-machine-learning-algorithms

29

Human-AI Interaction

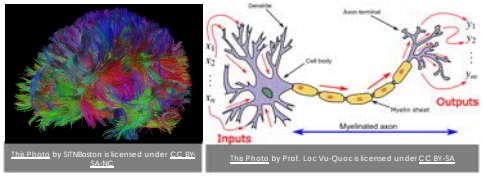
An emerging term...

- Human-AI Interaction (HAI/HAII)
- Human-Centred/Centric AI (HCAI)
- AI Interaction Design (AIxD)
- AI User Experience (AIUX)

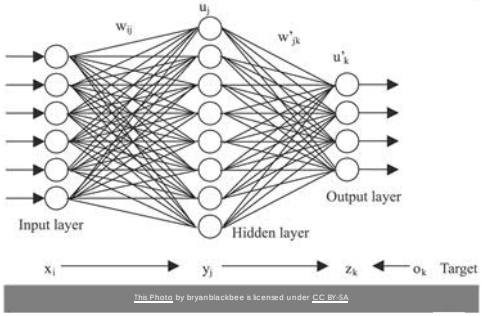
This Photo by Paul E. Baker is licensed under CC BY-NC-ND

30

Neural Networks



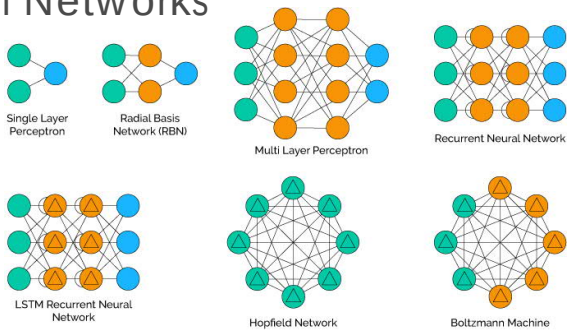
This Photo by SFNBoston is licensed under CC BY-SA-NC
This Photo by Prof. Loc Vu-Quoc is licensed under CC BY-SA



This Photo by bryanblackbee is licensed under CC BY-SA

31

Neural Networks

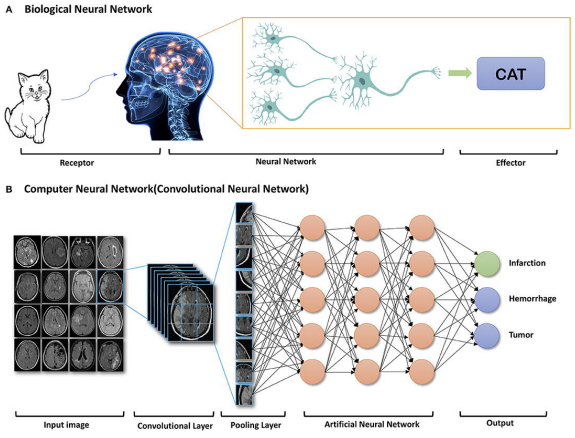


This Photo by Nikola M. Živković is licensed under CC BY

32

Deep Learning

- "Deep" = multiple layers (often 10+ "hidden layers")
- progressively extract higher-level features from the raw input
- Usually based on ANNs
- More in week 2



This Photo is licensed under CC BY.

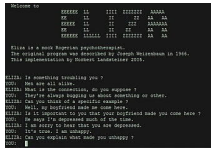
Zhu, G., Jiang, B., Tong, L., Xie, Y., Zaharchuk, G., & Wintermark, M. (2019). Applications of Deep Learning to Neuro-Imaging Techniques. *Frontiers in Neurology, 10*. https://doi.org/10.3389/fneur.2019.00869

33

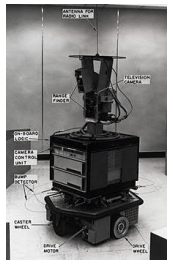
33

AI That Isn't Machine Learning

- Search-Based Approaches
  - A* search, for instance
  - Many early approaches for solving games
- Expert Systems
  - hand-coded rules and logic
  - "learn" by humans adding more rules
- (originally) Search Engines
  - early work focused on systems
- Rule/Grammar-Based NLP
  - But many projects move in grey zones



This Photo by Greg Jennings is licensed under CC BY-SA



34

This Photo by SRI International is licensed under CC BY-SA

34

Non-AI/ML Adaptive Systems

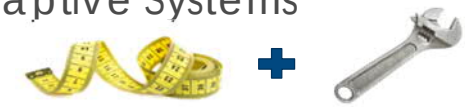
Focus on difficulty and input interpretation (parameters): > speed, accuracy + amplitude; parameter targeting range of motion of the player (e.g. reach and flexibility of arms)

(Smeddinck, Siegel & Herrlich, 2013. Adaptive Difficulty in Exergames for Parkinson's disease Patients. *Graphics Interface 2013.*)

35

35

Non-AI/ML Adaptive Systems

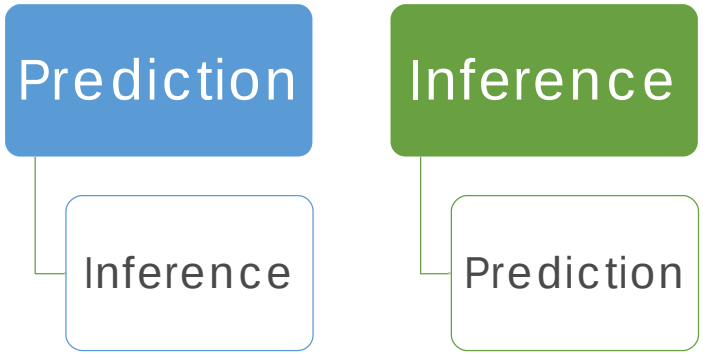


- Basic principle:
  - Can be heuristic or based on ML or user modeling
  - *performance evaluation + adjustment mechanism* (Adams, 2010)

36

36

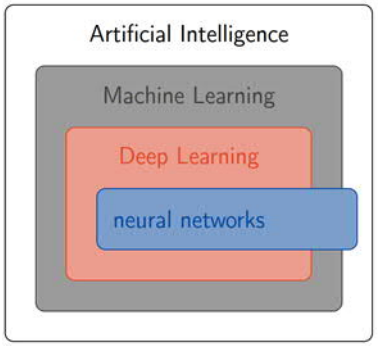
Machine Learning vs. Statistics



37

37

Quite the Hype



ca. 290.000 papers since 2015

ca. 630.000 papers since 2015

ca. 73.000 papers since 2015

ca. 300.000 papers since 2015

acc. to google scholar

Via Benjamin Paaßen, University of Syndey

38

38

Standard Machine Learning Pattern

"A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P if its performance at tasks in T, as measured by P, improves with experience E."
-- Tom M. Mitchell

Experience: Data (training / feedback)
Task: Prediction
(classification, recommendation, natural language generation, etc.)
Performance: Accuracy (and more)

Mitchell, T. M. (1997). *Machine Learning*. McGraw-Hill.

39

39

Standard ML Pattern

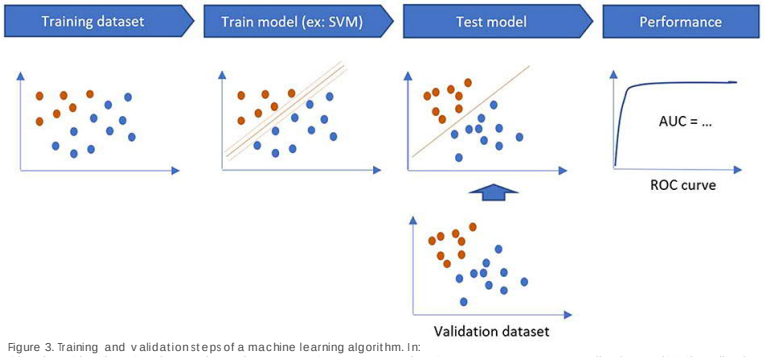
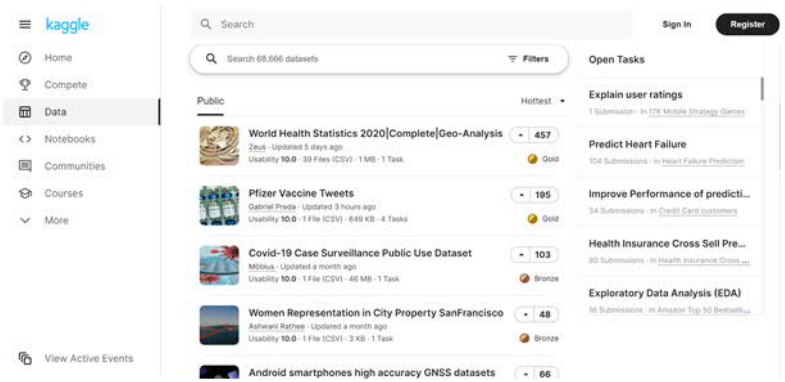


Figure 3. Training and validation steps of a machine learning algorithm. In:
Giraud, P., Giraud, P., Gasnier, A., El Ayachy, R., Kreps, S., Foy, J.-P., Durdux, C., Huguet, F., Burgun, A., & Bibault, J.-E. (2019). Radiomics and Machine Learning for Radiotherapy in Head and Neck Cancers. *Frontiers in Oncology*, *9*. https://doi.org/10.3389/fonc.2019.00174

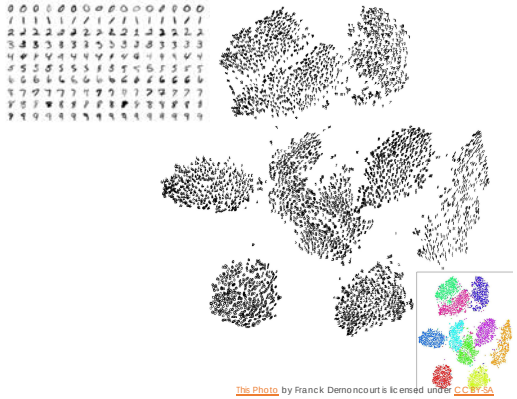
40

40

Open Data & Competitions



Famous Data Sets: MINST & IRIS



Fisher,R.A. "The use of multiple measurements in taxonomic problems" Annual Eugenics, 7, Part II, 179-188 (1936); also in "Contributions to Mathematical Statistics" (John Wiley, NY, 1950).

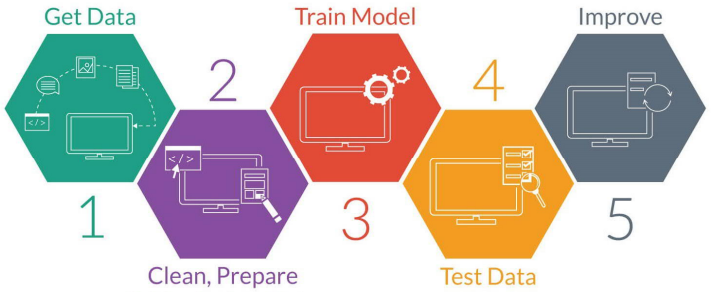
This Photo by Dlliff is licensed under CC BY-SA

Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
5.1	3.5	1.4	0.2	setosa
4.9	3.	1.4	0.2	setosa
4.7	3.2	1.3	0.2	setosa
4.6	3.1	1.5	0.2	setosa
5.	3.6	1.4	0.2	setosa
5.4	3.9	1.7	0.4	setosa
4.6	3.4	1.4	0.3	setosa
5.	3.4	1.5	0.2	setosa
4.4	2.9	1.4	0.2	setosa
4.9	3.1	1.5	0.1	setosa

This Photo by Franck Dernoncourt is licensed under CC BY-SA

This Photo by PhiChang is licensed under CC BY-SA

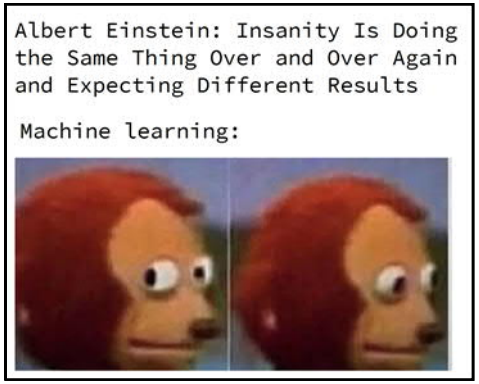
Machine Learning Lifecycle



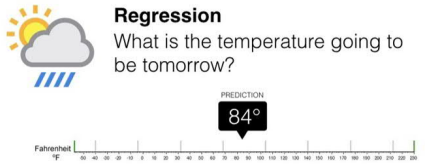
NOTICE SOMETHING?

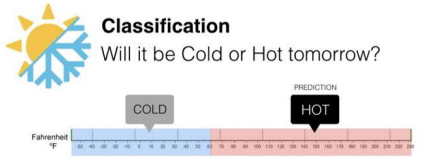
This Photo by Sunit Nandi / Techno FAQ is licensed under CC BY-SA-NC

Machine Learning Lifecycle

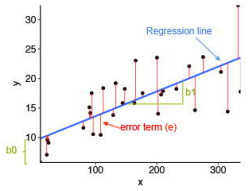


ML Tasks: Classification & Regression

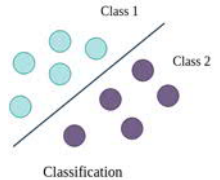




Via: https://colab.research.google.com/drive/1_g2w-LgVSyQjZiDqKaXYKyY7gVWLjbxvM



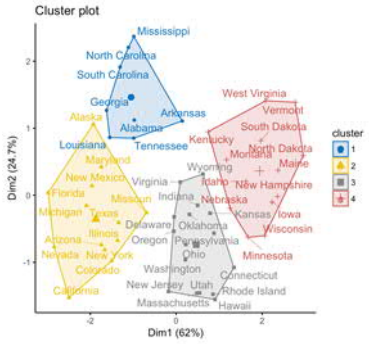
This Photo by kassambara is licensed under CC BY-SA-NC



This Photo by Anup Kumar and Bérénice Batut is licensed under CC BY

45

ML Tasks: Clustering (here: k-means; example of unsupervised learning)



This Photo by Alboukadel Kassambara is licensed under CC BY-SA-NC

46

Machine Learning Tasks

- Classification
- Regression
- Clustering
- Representation
- Task-Action Learning

- Map to Supervised, Unsupervised and Reinforcement Learning

47

Machine Learning Tasks

- Classification
  - Is this a dog or a cat?
- Regression
  - How warm will it be tomorrow?

Supervised learning

- Clustering
  - Here are a few articles. Organize them into topics
- Representation
  - What is the best way to represent words? (e.g. so representation of "autumn" and "fall" are similar)
  - What is the color of happiness?

Unsupervised learning

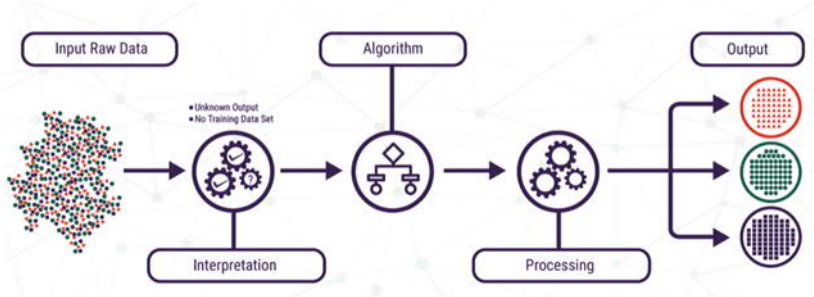
- "Dear algorithm: figure it out, and if you're right, you get a reward"
  - Drive a car
  - Recommend movies to people

Reinforcement learning

via https://haiicmu.github.io

48

Unsupervised Learning



This Photo by EP Labs is licensed under CC BY-SA

49

49

Supervised Learning

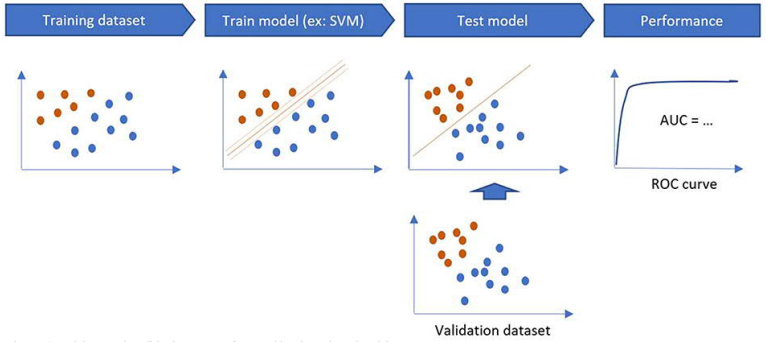
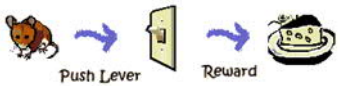


Figure 3. Training and validation steps of a machine learning algorithm. In: Giraud, P., Giraud, P., Gasnier, A., El Ayachy, R., Kreps, S., Foy, J.-P., Durdux, C., Huguet, F., Burgun, A., & Bibault, J.-E. (2019). Radiomics and Machine Learning for Radiotherapy in Head and Neck Cancers. *Frontiers in Oncology, 9*. https://doi.org/10.3389/fonc.2019.00174

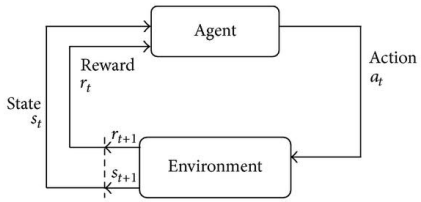
50

50

Reinforcement Learning



This Photo by Javier Benitez is licensed under CC BY-NC-ND



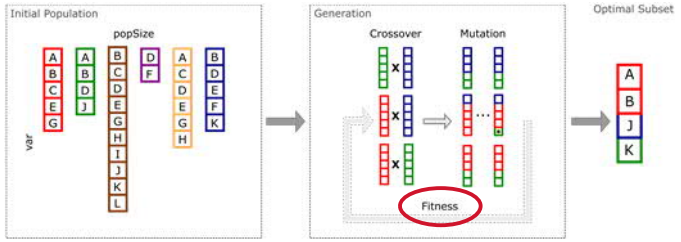
This Photo by Unknown Author is licensed under CC BY via Fu, Y., Liu, Q., Ling, X. & Cui, Z. (2014, January 28). *A Reward Optimization Method Based on Action Subrewards in Hierarchical Reinforcement Learning* [Research Article]. The Scientific World Journal; Hindawi. https://doi.org/10.1155/2014/120760

Heess, N., TB, D., Sriram, S., Lemmon, J., Merel, J., Wayne, G., Tassa, Y., Erez, T., Wang, Z., Eslami, S. M. A., Riedmiller, M., & Silver, D. (2017). Emergence of Locomotion Behaviours in Rich Environments. *ArXiv:1707.02286 [Cs]*. http://arxiv.org/abs/1707.02286

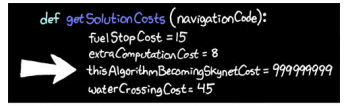
51

51

Genetic Algorithms



This Photo by Gustavo is licensed under CC BY-SA



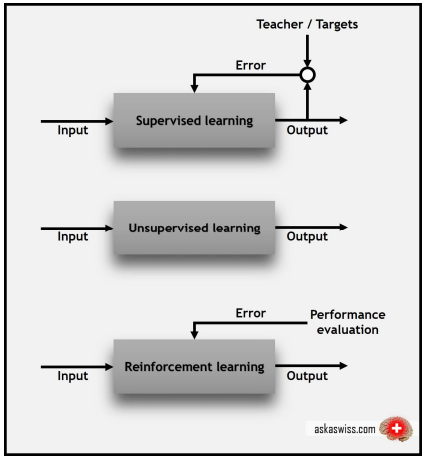
GENETIC ALGORITHMS TIP:
ALWAYS INCLUDE THIS IN YOUR FITNESS FUNCTION

This Photo by Joshua Eckroth is licensed under CC BY-SA

52

52

Learning Principles Overview



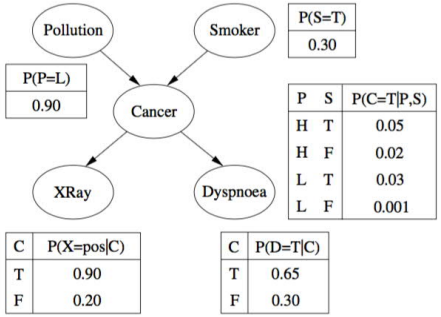
Michael Beyeler CC-BY

Complex / Mixed Methods

- Ensembles
- Auto-ML
- Cognitive Frameworks

Bayesian Networks

Pearl J (1985). Bayesian Networks: A Model of Self-Activated Memory for Evidential Reasoning (UCLA Technical Report CSD-850017). Proceedings of the 7th Conference of the Cognitive Science Society, University of California, Irvine, CA. pp. 329–334. Retrieved 2009-05-01.



P(S=T)
0.30

P(P=L)
0.90

P	S	P(C=T|P,S)
H	T	0.05
H	F	0.02
L	T	0.03
L	F	0.001

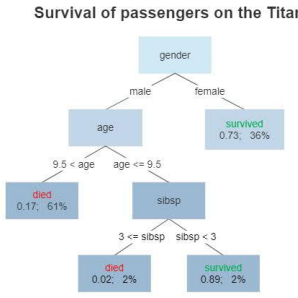
C	P(X=pos|C)
T	0.90
F	0.20

C	P(D=T|C)
T	0.65
F	0.30

FIGURE 2.1
A BN for the lung cancer problem.

This Photo by Franck Dernoncourt is licensed under CC BY-SA

Prominent Algorithms: Decision Trees



CC-BY: Gilgoldm

- Classification trees
- Regression trees
- Multiple tree ensembles: e.g. random forest
- Simple to understand and interpret
- Learning optimal trees is NP-complete
- Trees can be over-complex

Prominent Algorithms Summary

- NN
- Regression analysis
- K-means
- Genetic algorithms
- Bayesian networks
- Decision trees

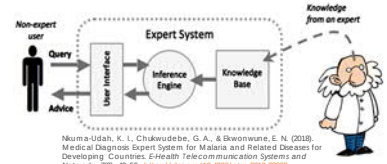
- ... more (and more details) in week 2!

57

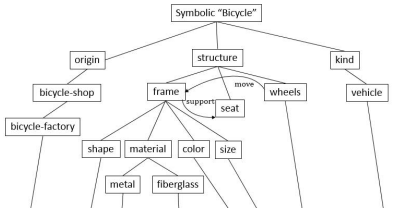
57

Symbolic AI

E.g. Expert Systems



Nkuma-Udah, K. I., Chukwudebe, G. A., & Ekwonwune, E. N. (2018). Medical Diagnosis Expert System for Malaria and Related Diseases for Developing Countries. *E-Health Telecommunication Systems and Networks*, *7*(2), 43-56. https://doi.org/10.4236/etsn.2018.72005



This Photo by Duke SciPols licensed under CC BY-SA

This Photo by Michael L. Umbricht and Carl R. Friend is licensed under CC BY-SA

58

58

Applications: Natural Language Processing

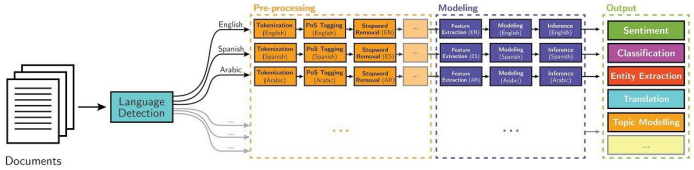
- Rich history ... may or may not make use of AI/ML
- NLP: NLU (understanding) or NLG (generation)
- Many applications:
  - Sentiment Analysis
  - Machine Translation
  - Question Answering
  - Text Summarization
  - Text Classification
  - Text-to-Speech
  - Speech Recognition

Intro at:
https://devopedia.org/natural-language-processing

59

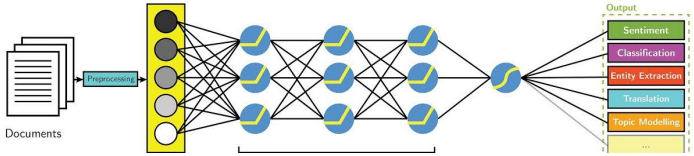
59

Applications: Natural Language Processing



Classical NLP (often symbolic) has given way to Deep Learning NLP. Source: Le 2018.

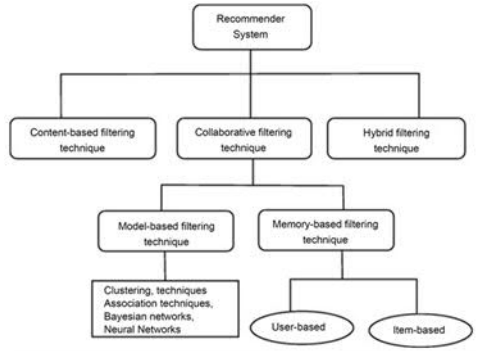
Deep Learning-based NLP



60

60

Applications: Recommender Systems



Mylavarapu, B. K. (2018). Collaborative Filtering and Artificial Neural Network Based Recommendation System for Advanced Applications. *Journal of Computer and Communications*, 6(12), 1–14. https://doi.org/10.4236/jcc.2018.612001

61

Applications: Recommender Systems

- Collaborative filtering

	i_1	i_2	i_3	i_4	i_5	...	i_M
Jack	5	1	3	4	3	...	1
u_2	0	4	0	0	0	...	5
u_3	0	2	0	0	0	...	4
Tom	4	1	4	5	4	...	2
...	...	...	...	...	...	...	...
Myself	5	1	4	?	3	...	1

similar users based on profile

target user

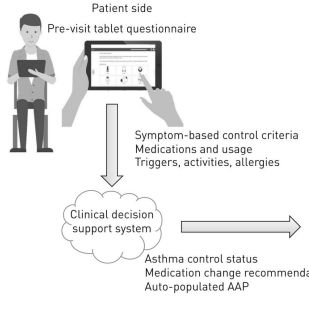
predicted rating: 4.5

This Photo by WikiBooks is licensed under CC BY-SA

62

Applications: Recommender Systems

- Decision support

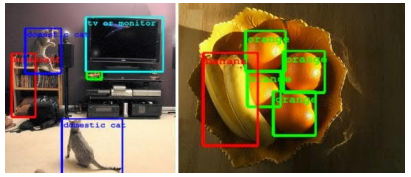


This Photo by Enigma3542002 is licensed under CC BY-SA

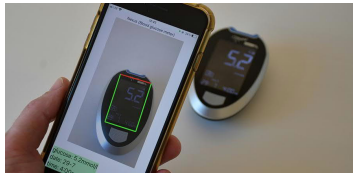
Gupta, S., Price, C., Agarwal, G., Chan, D., Goel, S., Boulet, L.-P., Kaplan, A. G., Lebovic, G., Mamdani, M., & Straus, S. E. (2019). The Electronic Asthma Management System (eAMS) improves primary care asthma management. *European Respiratory Journal*, *53*(4).

63

Applications: Computer Vision



This Photo by Alex Chitu is licensed under CC BY



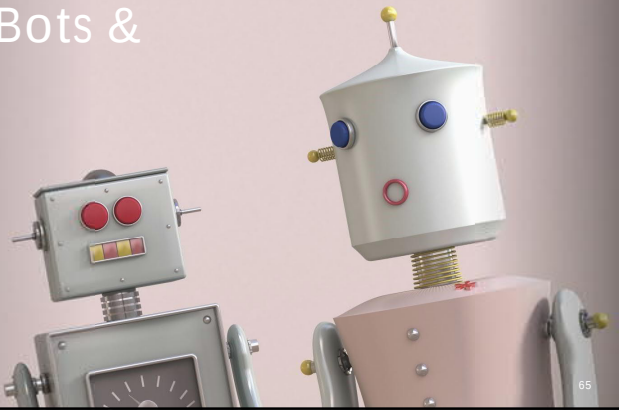
This Photo by Alex Chitu is licensed under CC BY

Dean, J., Corrado, G. S., Monga, R., Chen, K., Devin, M., Le, Q. V., Mao, M. Z., Ranzato, M., Senior, A., Tucker, P., Yang, K., & Ng, A. Y. (2012). Large Scale Distributed Deep Networks. *NIPS*.

James Charles, Stefano Bucciarelli and Roberto Cipolla. 'Real-time screen reading: reducing domain shift for one-shot learning.' Paper presented at the British Machine Vision Conference. 2020

64

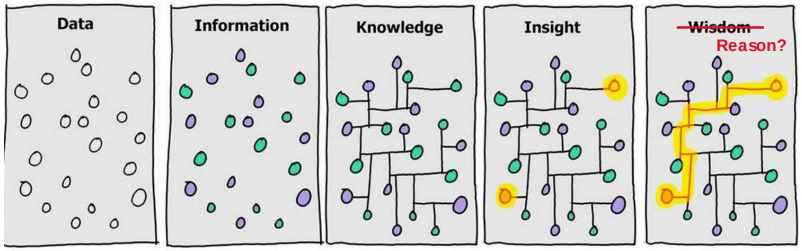
Applications: Agents, Bots & Robots



65

65

Artificial General Intelligence?



66

66

Ubiquitous Computing → AI

- Dynamic / learning / adaptive systems becoming ubiquitous (esp. online services / mobile devices)
- AI/ML also becoming ubiquitous
- Commodity
- Still centralised (esp. amongst FAANG)
- Computing as “tools” to “services” and even “servants”
- “There’s an app for that” → “There’s an agent for that”

67

67

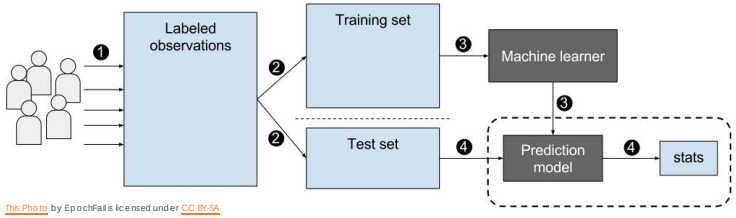
BREAK

CC-BY Mind Data AI

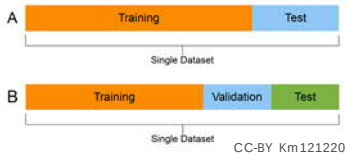
68

68

Models + Training & Evaluation Data



This Photo by EpochFail is licensed under CC BY-SA



Dataset A: test set used to test trained model
Dataset B: validation set used to test trained model, test set to evaluate final model

CC-BY Km121220

Evaluating Learning Systems

Generally interested in:

- How often is the prediction wrong?
- How is the prediction wrong?
- What is the cost of wrong predictions?
- How does the cost vary by the type of prediction that was wrong?
- How can we minimize cost?

- *HCAI: misunderstandings? frustration? regret?*

Accuracy

correct predictions
---
total predictions

This Photo by Zabaware is licensed under CC BY-ND

Accuracy

From the Guardian: https://www.theguardian.com/us-news/2016/dec/14/dnc-hilary-clinton-emails-hacked-russia-aide-typo-investigation-finds and NYT: https://www.nytimes.com/2016/12/13/us/politics/russia-hack-election-dnc.html?hp&action=click&pgtype=Homepage&clickSource=story-heading&module=a-lede-package-region&region=top-news&WT.nav=top-news

via https://haiicmu.github.io

Measuring Accuracy (Classes)

False Positive and *False Negative* Conclusions

Actual (**True**) Condition → ↓ Evaluation Conclusion	Evaluee *has* claimed mental disorder	Evaluee does *not* have claimed mental disorder
Evaluee *has* claimed mental disorder	True Positive	*False Positive*
Evaluee does *not* have claimed mental disorder	*False Negative*	True Negative

This Photo by Unknown Author is licensed under CC BY-SA

Weigh the impacts as an important planning step in human-AI interaction!

73

73

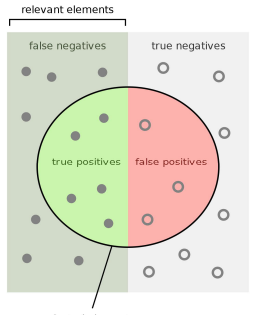
Sensitivity & Specicifity

Sensitivity == True Positive Rate

- e.g. ability of the test to identify those who have a disease

Specificity == True Negative Rate

- e.g. ability of the test to identify those without the disease



This Photo by FeanDoe is licensed under CC BY-SA

74

74

FALSE POSITIVES: SELF-DRIVING CARS AND THE AGONY OF KNOWING WHAT MATTERS

via https://haiicmu.github.io

According to a preliminary report released by the National Transportation Safety Board last week, Uber's system detected pedestrian Elaine Herzberg six seconds before striking and killing her. It identified her as an unknown object, then a vehicle, then finally a bicycle. (She was pushing a bike, so close enough.) About a second before the crash, the system determined it needed to slam on the brakes. But Uber hadn't set up its system to act on that decision, the NTSB explained in the report. The engineers prevented their car from making that call on its own "to reduce the potential for erratic vehicle behavior." (The company relied on the car's human operator to avoid crashes, which is a whole separate problem.)

ARN MORE

Uber's engineers decided not to let the car auto-brake because they were worried the system would overreact to things that were unimportant or not there at all. They were, in other words, very worried about false positives.

https://www.wired.com/story/self-driving-cars-uber-crash-false-positive-negative/

75

75

Precision & Recall

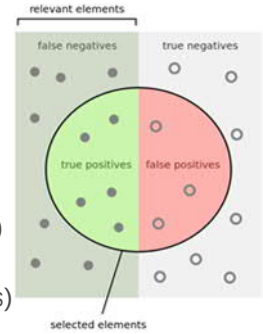
Establish relations between positives and negatives...

Precision =
true positives / (true positives + false positives)

Recall =
true positives / (true positives + false negatives)

Notice something?

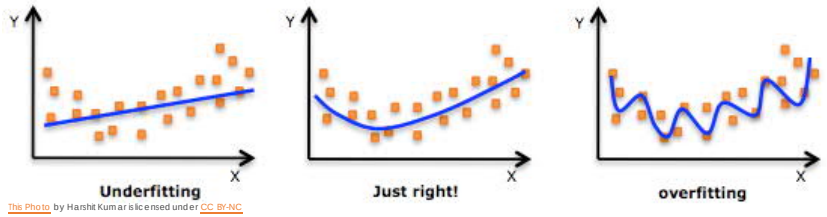
Recall == Sensitivity



76

76

Overfitting, Underfitting



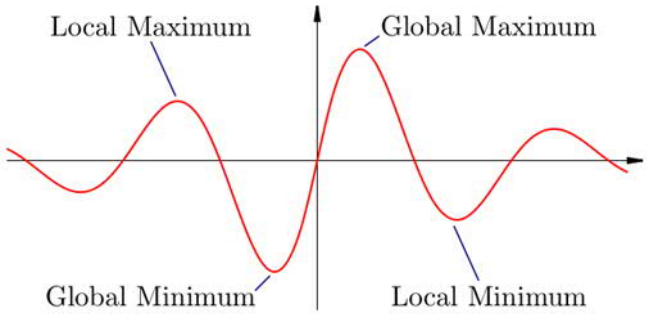
This Photo by Harshit Kumar is licensed under CC BY-NC

underfitting → too many errors in sample data
overfitting → too many errors on unseen data in deployment

77

77

Local Minima / Maxima



78

78

Machine Learning A Cappella - Overfitting Thriller!

https://www.youtube.com/watch?v=DQWI1kvmwRg

79

79

Silent Errors? UX?

- Not all errors are visible to the computer
- How do you measure when your device can't hear "OK, Google"
- More in later sessions…

80

80

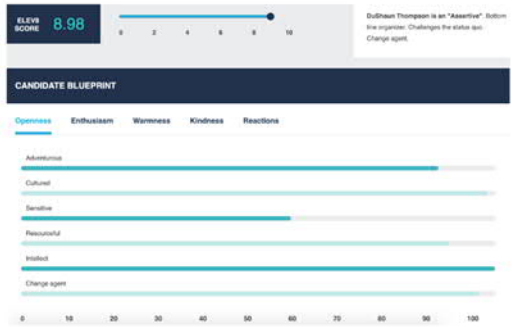
Biases & Fairness

- Bias as in prejudice in favor or against a person, group, or thing that is considered to be unfair.
- Frequently see bias in the classifications and predictions
- Often due to how data is sampled / collected
- Can also be introduced in processing
- Many algorithms/methods now in development to analyse / check for biases / fairness
- Proper judgment may require humans-in-the-loop (more in week 3)

81

81

Fair? Accurate?



Raghavan, M., Barocas, S., Kleinberg, J., & Levy, K. (2020). Mitigating Bias in Algorithmic Hiring: Evaluating Claims and Practices. *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, 469–481. https://doi.org/10.1145/3351095.3372828

82

82

AI Snake Oil

Vendor name	Funding	# of employees	Location
8 and Above	–	1-10	WA, USA
ActiView	$6.5M	11-50	Israel
Applied	£2M	11-50	UK
Assessment Innovation	$1.3M	1-10	NY, USA
Good&Co	$10.3M	51-100	CA, USA
Harver	$14M	51-100	NY, USA
HireVue	$93M	251-500	UT, USA
impress.ai	$1.4M	11-50	Singapore
Knockri	–	11-50	Canada
Koru	$15.6M	11-50	WA, USA
LaunchPad Recruits	£2M	11-50	UK
myInterview	$1.4M	1-10	Australia
Plum.io	$1.9M	11-50	Canada
PredictiveHire	A$4.3M	11-50	Australia
pymetrics	$56.6M	51-100	NY, USA
Scoutible	$6.5M	1-10	CA, USA
Teamscope	€800K	1-10	Estonia
ThriveMap	£781K	1-10	UK
Yobs	$1M	11-50	CA, USA

Not all these companies offer AI assessments of job candidates, but most do.

Raghavan et al. *Mitigating Bias in Algorithmic Hiring: Evaluating Claims and Practices*. Preprint 2019.

83

83

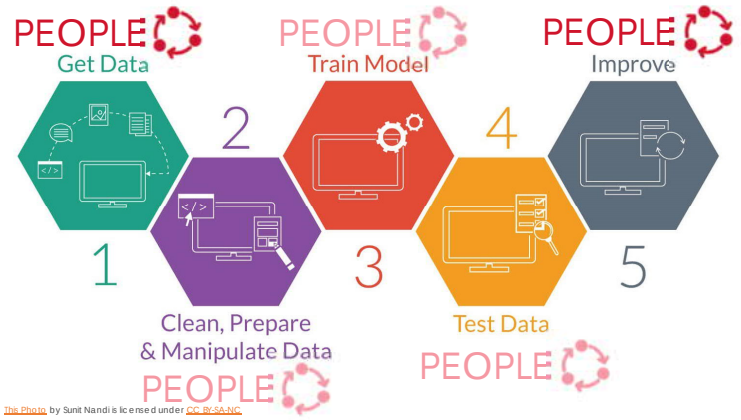
Rough Compass on AI Capabilities

- Perception: rapid, legitimate process
- Automating judgment (spam detection, essay grading, detecting hate speech, content recommendations, etc.): measurable progress but stark imperfections
- Predicting social outcomes (e.g. job success, policing, terrorist risk, etc.): VERY QUESTIONNABLE!
  - In most cases: manual scoring rules as accurate and more transparent

84

84

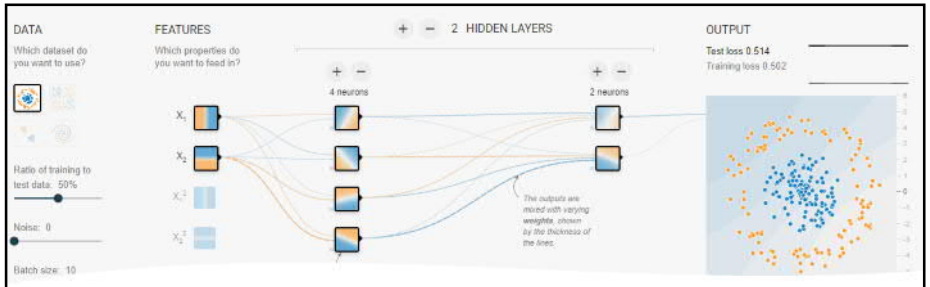
Human-AI Interaction: Preview



This Photo by Sunit Nandi is licensed under CC BY-SA-NC

85

85



Neural Network Playground

- Put those concepts to use (and learn a few more terms that you'd otherwise hear about in week 2)...
- https://playground.tensorflow.org/

86

86

Human-AI Interaction & Futures

Week 01 – Session 03:
AI History & Perspectives

Jan Smeddinck

Unless otherwise noted, all materials: (CC BY-NC-SA 4.0)

87

87

Learning Goals

- Rough history of AI...
- ...with bits of history of HCI

Timeline:

88

88

Hype Cycles

- AI spring: time period when new technology prompts a period of intense funding and research around AI
- Breaking point: where expectations fail to meet reality
- AI winter: time period with lower funding given to AI; research is slower or more incremental

This Photo by Jeremykemp is licensed under CC BY-SA

89

89

Don't Promise What You Can't Keep

"Machines will be capable, within twenty years, of doing any work a man can do."

- Herbert Simon 1965

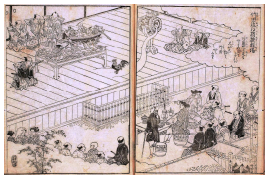
"In from three to eight years we will have a machine with the general intelligence of an average human being."

- Marvin Minsky 1970

90

90

Pygmalion priant Vénus d'animer sa statue, Jean-Baptiste Regnault



Automaton theatre circa 1662, Osaka Japan

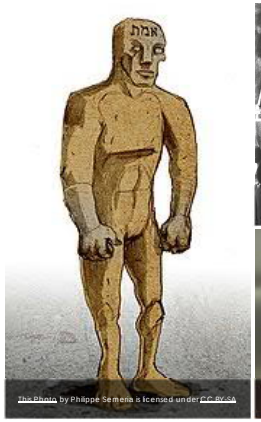
The flute-player by Innocenzo Manzetti (1840)

Ancient History of AI

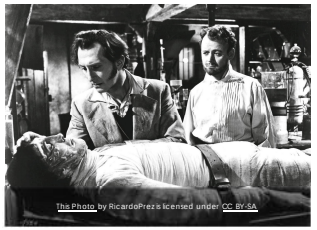
91

91

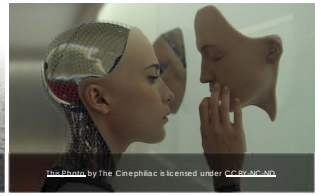
Golem, Frankenstein (1818), Ex-Machina



This Photo by Philippe Semeria is licensed under CC BY-SA



This Photo by RicardoPrez is licensed under CC BY-SA



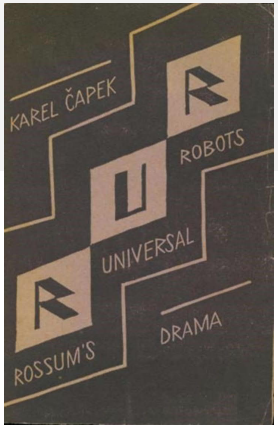
This Photo by The Cinephiliac is licensed under CC BY-NC-ND

92

92

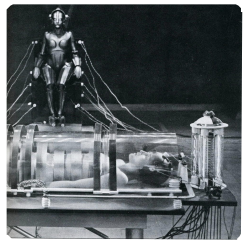
Robots (1920)

- Slavic: robota ~ forced labor
- 1920 play:
  Rossumovi Univerzální Roboti
  Rossum's Universal Robots
- Czech Karel Čapek
- Fictional humanoids



93

93



Metropolis (1927)

- 1927, Director: Fritz Lang
- ~ 1st feature length Science Fiction film
- Written by Thea V on Harbou
- Early Bladerunner (stylish)
- life in 2029 (overpasses, mobile / video phones, video surveillance, TV, etc.)
- And of course C3PO
- Initially censored … search for long form if interested in watching

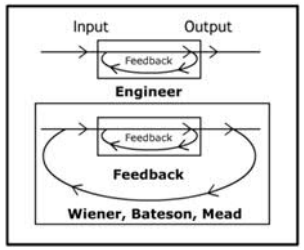
https://archive.org/search.php?query=metropolis&and%5B%5D=mediatype%3A%22movies%22&sort=-week

94

94

Cybernetics (1940ish)

- Transdisciplinary approach for exploring regulatory / purposive systems
  - Structures, constraints, possibilities
  - Early ideas about embodiment (more later)
  - Process-oriented: designing, learning, managing, conversing
- Interdicsiplinary Macy Conferences (1941 – 1960)
- "the scientific study of control and communication in the animal and the machine"
  - Wiener, Norbert (1948). Cybernetics: Or Control and Communication in the Animal and the Machine. Cambridge, Massachusetts: MITPress.
- Roots of AI (split in 1950s)
- Roots of HCI



Second-order cybernetics … recursive application …

This Photo by Clockwork is licensed under CC BY-SA

95

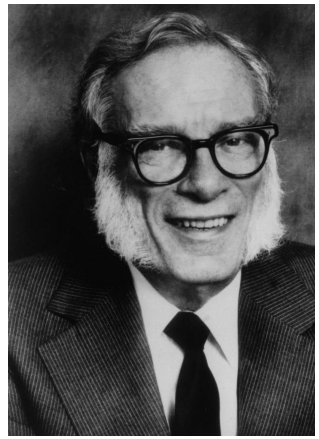
95

Robots #2 (1942)

1942 Isaac Asimov's Robot Laws

1. A robot may not injure a human being or, through inaction, allow a human being to come to harm.
2. A robot must obey the orders given it by human beings except where such orders would conflict with the First Law.
3. A robot must protect its own existence as long as such protection does not conflict with the First or Second Laws.

- More in later sessions…



This Photo by Big Jack is licensed under CC BY-SA

96

96

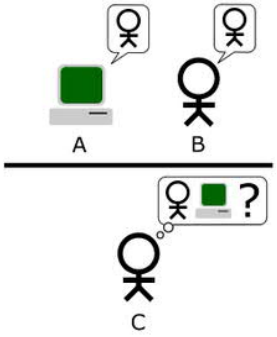
Turing Test (1950)

Lots of people: What truly *is* human intelligence?

Alan Turing: How can we decide when a machine has achieved human-level intelligence?

→ imitation game aka Turing Test

MUST READ: Turing, A. M. (1950). Computing machinery and intelligence. *Mind*, *59*(236), 433–460.



This Photo by Hugo Férée is licensed under CC BY-SA

97

97

Dartmouth in 1956

- Dartmouth Summer Research Project on Artificial Intelligence
- AI starts to form as a field from roots in cybernetics, automata theory, complex systems, etc.
- Foundational all-stars of AI: Marvin Minsky, Julian Bigelow, D.M. Mackay, Ray Solomonoff, John Holland, John McCarthy, Claude Shannon, Nathanial Rochester, Oliver Selfridge, Allen Newell, Herbert Simon

Marvin Minsky, MIT

John McCarthy, Dartmouth/ Stanford

Herbert Simon, CMU

Allan Newell, CMU

98

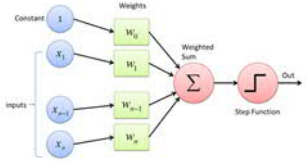
98

Connectionism & Perceptron (1958)

- The Perceptron: designed by Frank Rosenblatt 1958
- To run on Mark 1 Perceptron machine
- US Navy image recognition, linking 400 photocells

"the embryo of an electronic computer that [the Navy] expects will be able to walk, talk, see, write, reproduce itself and be conscious of its existence." - NYT 1958 (following comments by Rosenblatt)

$$f(\mathbf{x}) = \begin{cases} 1 & \text{if } \mathbf{w} \cdot \mathbf{x} + b > 0, \\ 0 & \text{otherwise} \end{cases}$$



This Photo by SAGAR SHARMA is licensed under CC BY-SA

99

99

Skinner teaching machine (1960)

- Roots of modern "e-learning" …

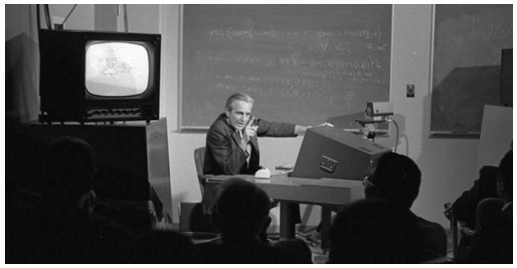
https://www.youtube.com/watch?v=CFYruzWeFwQ

100

100

Engelbart (1968) (and McLuhan)

- "Mother of all demos" ... early HCI ...
- "A Research Centre for Augmenting Human Intellect"



https://www.youtube.com/watch?v=Xptc6f3Daoo

101

101

AI vs. IA

Artificial intelligence ~ building a human-like intelligence (for specific kinds of tasks?)

Intelligent augmentation/amplification:

- Computing machines will do routine work; prepare the way for insights and decisions in technical and scientific thinking (Licklider)
- "elegant combination of reasoning machinery and direct manipulation" (Horvitz)

102

1969 "Perceptron" (destroyed)

- Marvin Minsky and Seymour Papert publish a book "Perceptrons"
- Burn piece of Perceptron approach in favour of rule approach
- Perceptron cannot handle XOR operator
  - However, multi-layer perceptrons CAN (and this was known at the time)
- Shuts down funding for neural networks
- Rosenblatt soon dies, never to see neural nets revindicated

103

103

Symbolic AI

- Symbolism: formal logic systems can represent intelligent action
- Assumption: "*Every aspect of learning or any other feature of intelligence can be so precisely described that a machine can be made to simulate it.*"
- Later to be enshrined in Lisp-Machines

```
(RULE 5
     (IF (PCS-SCS HEAT TRANSFER INADEQUATE)
         (LOW FEEDWATER FLOW))
     (THEN (ACCIDENT IS LOSS OF FEEDWATER)))

(RULE 6
     (IF (SG INVENTORY INADEQUATE)
         (LOW FEEDWATER FLOW))
     (THEN (ACCIDENT IS LOSS OF FEEDWATER)))

(RULE 7
     (IF (PCS INTEGRITY CHALLENGED)
         (CONTAINMENT INTEGRITY CHALLENGED))
     (THEN (ACCIDENT IS LOCA)))

(RULE 8
     (IF (PCS INTEGRITY CHALLENGED)
         (SG LEVEL INCREASING))
     (THEN (ACCIDENT IS STEAM GENERATOR TUBE
     RUPTURE)))

(RULE 9
     (IF (SG INVENTORY INADEQUATE)
         (HIGH STEAM FLOW))
     THEN (ACCIDENT IS STEAM LINE BREAK)))
```

Figure 2. Event-oriented IF-THEN rules.

104

104

Symbolic AI

- Newell & Simon's *General Problem Solver* can solve math proofs by searching a logic space
- Advances in natural language processing based on rules how words relate
- Advances in computer vision based on image transforms
- Advances in robotics based on rules and search in simplified settings
- *BUT: Strong limitations in scalability!*

105

105

1st AI Winter (~1974)

Funding is lost due to unmet promises:

- Lighthill Report 1973 shuts down funding in UK
  - James Lighthill (1973): "Artificial Intelligence: A General Survey" in Artificial Intelligence: a paper symposium, Science Research Council
- DARPA switches to "mission-oriented direct research, rather than basic undirected research"
- Dreyfus at MIT argues lots of human reasoning is *not based on logic rules*, involving instinct and unconscious reasoning
  - (No AI researcher will eat lunch with Dreyfus for the next decade)
- Sussman: "*using precise language to describe essentially imprecise concepts doesn't make them any more precise*."

106

106

Limitations with Logic rules: Combinatorial Explosion

- Representation of knowledge is hard
- All the possibilities (mutability vs. immutability of the sign … semioticians such as de Saussure, Charles Peirce et al.)
- Combinatorial explosion and search space
- Common sense

- Robot motion planning fails: https://www.youtube.com/watch?v=g0TaYhjpOfo

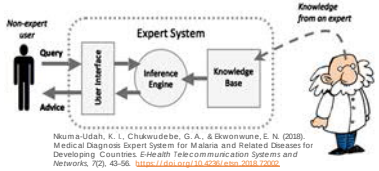
107

107

1980's AI Spring: From General Intelligence to Specific Applications

- Expert systems: while rules can't do everything they are useful for capturing what experts do
- 1989 Deep Thought at CMU can play human-level chess

Expert Systems (great prototypes, but riddled by scalability issues)



Nkuma-Udah, K. I., Chukwudebe, G. A., & Ekwonwune, E. N. (2018). Medical Diagnosis Expert System for Malaria and Related Diseases for Developing Countries. *E-Health Telecommunication Systems and Networks, 7*(2), 43-56. https://doi.org/10.4236/etsn.2018.72002

This Photo by Michael L. Umbricht and Carl R. Friend is licensed under CC BY-SA

- *Again: hype gets too high and funding is lost! → 2nd AI winter (into 90s / 2000s … depending on perspective)*

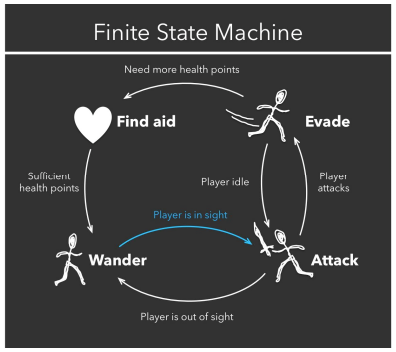
108

108

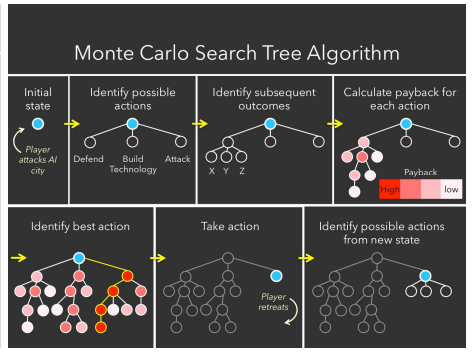
1990s - 2000s AI "undercover"

- Many advances in computer science, probability theory, statistical learning, etc.
- Video games boom births GPUs...
- *"Computer scientists and software engineers avoided the term artificial intelligence for fear of being viewed as wild-eyed dreamers."* - NYT 2005
- Machine Learning as a field distances itself from AI (temporarily!)
- Online search and advertising push for personalisation / recommendations / user models
- General public "goes online" ... lots of data becomes available
- Structured approaches in "linked data" / ontologies

Video Games & AI (Symbiotic!?)



This Photo by Shannon McArdel is licensed under CC BY-SA-NC

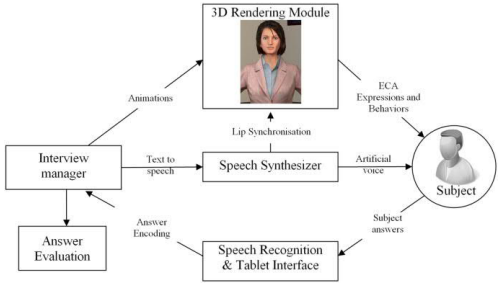


This Photo by Shannon McArdel is licensed under CC BY-SA-NC

http://sitn.hms.harvard.edu/flash/2017/ai-video-games-toward-intelligent-game/

ECAs & NPCs

Embodied Conversational Agents & Non-Player Characters (in video games)



Philip, P., Micoulaud-Franchi, J.-A., Sagaspe, P., Sevin, E. D., Olive, J., Bioulac, S., & Sauteraud, A. (2017). Virtual human as a new diagnostic tool, a proof of concept study in the field of major depressive disorders. *Scientific Reports*, 7(1), 42656. https://doi.org/10.1038/srep42656

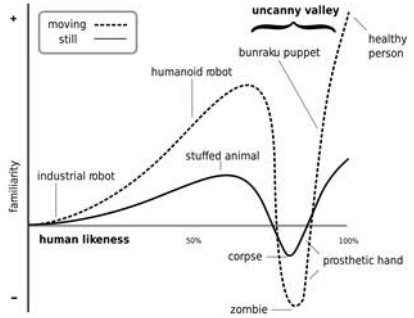
Becker, C., Kopp, S., & Wachsmuth, I. (2004). Simulating the Emotion Dynamics of a Multimodal Conversational Agent. In *Affective Dialogue Systems* (pp. 154–165). http://www.springerlink.com/content/1d5493hfwedq083f

Uncanny Valley

Mori, M. (1970). Bukimi no tani [The uncanny valley]. Energy, 7(4), 33–35.

This Photo by Max Braun is licensed under CC BY-SA

Replicant Q2 T.Minato, T. Shimada, M. Ishiguro, H. Itakura & Kokoro



This Photo by Unknown Author is licensed under CC BY-SA

History of Human-AI Interaction

2 Reads:

- Shneiderman, B., & Maes, P. (1997). Direct manipulation vs. Interface agents. *Interactions*, *4*(6), 42–61. https://doi.org/10.1145/267505.267514
- Horvitz, E. (1999). Principles of mixed-initiative user interfaces. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems the CHI Is the Limit - CHI '99*, 159–166. https://doi.org/10.1145/302979.303030
  - Akin to Shneiderman's Golden Rules / Nielsen's Heuristics in UB

113

113

History of Human-AI Interaction

Horvitz, E. (1999). Principles of mixed-initiative user interfaces. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems the CHI Is the Limit - CHI '99*, 159–166. https://doi.org/10.1145/302979.303030

- Akin to Shneiderman's Golden Rules / Nielsen's Heuristics in UB

Critical factors for the effective integration of automated services with direct manipulation interfaces include:

1. Developing significant value-added automation
2. Considering uncertainty about a user's goals
3. Considering the status of a user's attention in the timing of services
4. …

114

114

Machine Learning Boom (2010ish)

- By 2010s reached serious scale
- Moore's law + data, data, data
- Machine learning now immensely powerful (perception)
- Neural networks now practical, many earlier inventions in neural networks (e.g. back propagation from 1980s) are "re-discovered"
- Deep learning gives big performance leaps in almost all application areas overnight
- Massive funding is back (now combined research + "real-world")
- ML "re-unites" with AI … starts to become pervasive / ubiquitous
- *HCI has since gone from UB → UX … with widespread use, neglect of user-centred approaches in learning systems becomes obvious*

115

115

Timeline Recap

116

116

Human-AI Interaction & Futures

Week 01 – Session 04:
Human-Data Interaction

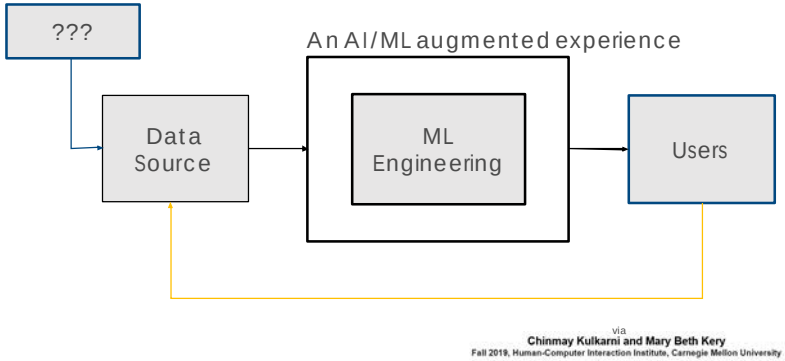
Jan Smeddinck

Unless otherwise noted,
all materials: (CC BY-NC-SA 4.0)

117

117

Where does data come from?



via
Chinmay Kulkarni and Mary Beth Kery
Fall 2019, Human-Computer Interaction Institute, Carnegie Mellon University

118

118

Standard Open Data Sources

- Traditional methods / algorithms development in AI frequently uses existing data sets (e.g. IRIS, MINST, ImageNet, etc.) ... i.e. published research data
  - Readily available; often well formatted / relatively complete
  - Good for benchmarking
- Cost / complexity of acquiring / processing / managing data often overlooked

- *Applied AI/ML often needs real-world relevant data sets and builds on „user data“...*
  - *Can be difficult to get by, unorganised, sparse, legal implications*
  - *Difficult to share, hence difficult to compare*

119

119

Questions About Open Data

Usman Haque (building on Rittel and Webber's 1973: wicked problems in social policy planning; along example of "smart cities"):

- *assumes if we had enough data we could make perfect decisions*
- *attempt to understand, explain, control*
  - *"It's not me, it's the data, which is impartial"*
- *Borges metaphors: map, library, encyclopedia*
  - *implies infinite data, merely need an index and our problems will be solved*
- *often only the least controversial data is 'opened'*
  - *transportation data, not wikileaks, etc.*
- *is 'open data' just letting off steam?*
- *city is a mess. not infrastructure. the algorithm god isn't going to solve this.*
- *embrace super wicked urban problems, don't reverse-engineer problems based on existing solutions*

120

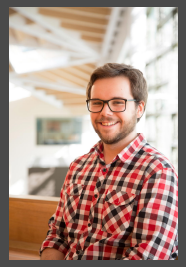
120

Justification Practices

Open Lab

From ongoing work by Kieran Cutting (supervisee):

- Design under austerity
- Work with young people in state foster care at transition to adulthood
- HCI now embraces postcolonialism, feminism, and anarchism
- Little associated methodological innovation
- Researchers to rely upon old methods in new contexts
- Justification practices: grounded theory how work, relationships, and experiences of care have changed as a result of austerity
- Can subsume well-intended technology design and use



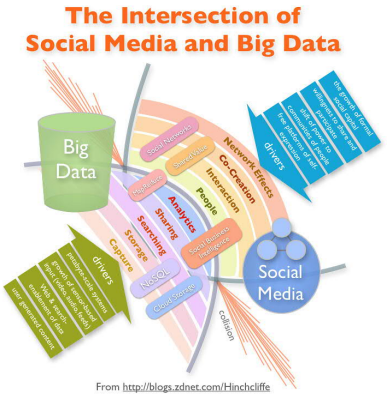
K.L.Cutting2@newcastle.ac.uk

121

121

Search & Stats, Advertising, Social Media

- Historically as enabled by advent of (broad) internet use
- Also public data (beyond pure research data) under "open data" initiatives…



From http://blogs.zdnet.com/Hinchcliffe

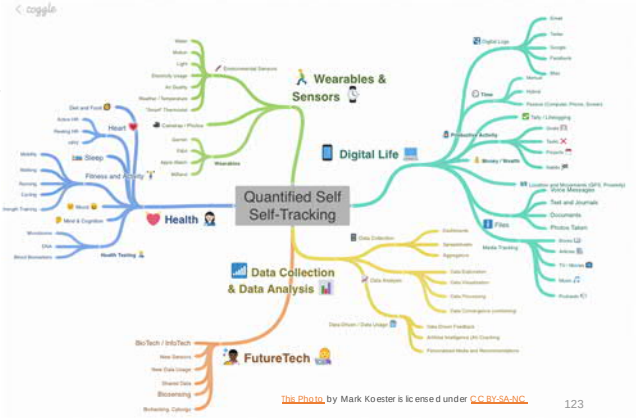
This Photo by Laura James is licensed under CC BY
This Photo by Dion Hinchcliffe is licensed under CC BY-SA

122

122

Quantified Self & Personal Informatics

- Innovating communities of early-adopters / avant-garde
- „self-surveillance"
- Nth senses
- Links to IoT
- History since 1970s



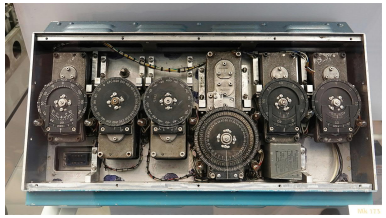
This Photo by Mark Koester is licensed under CC BY-SA-NC

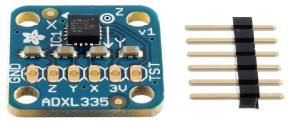
123

123

Sensing Devices (Wearables & Other)

- Rapid development in sensor technologies … e.g. accelerometers in 1950s ~ $100k+ … today: cents





This Photo by Sanjay Acharya is licensed under CC BY-SA
This Photo by oomlout is licensed under CC BY-SA

124

124

Sensing Devices (Wearables & Other)

- Smart homes, ambient assisted living, etc.
- Smart cities (Newcastle, USB)
  - https://newcastle.urbanobservatory.ac.uk/
  - 1,423,620,797 records (as of 2021-01-26)
  - https://3d.usb.urbanobservatory.ac.uk/
  - https://api.usb.urbanobservatory.ac.uk/live/
- active vs. passive data collection

125

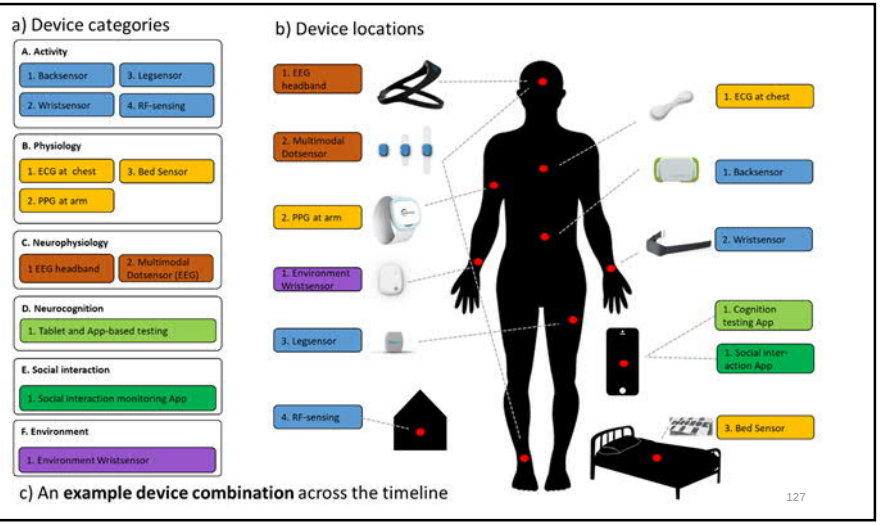
125

IDEA FAST

Identifying Digital Endpoints to Assess FAtigue, Sleep and acTivities of daily living in neurodegenerative disorders and immune-mediated inflammatory diseases

IDEA-FAST 853981 proposal, IMI2 15th Call, topic 6

126

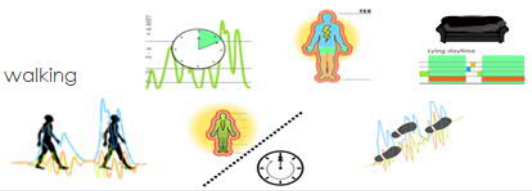


127

127

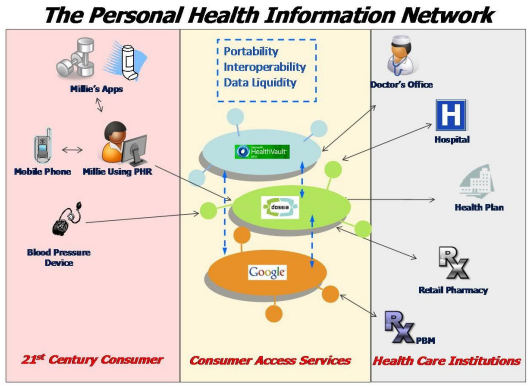
What are we going to measure?

- Exact combination of devices will be determined in the first 3 months of IDEA-FAST (to include recent developments and allow in-depth literature review)
- Example 1: Dreem headband for sleep
  - Dry EEG sensors (3 frontal, 2 occipital)
  - Pulse oximeter (heart-rate)
  - 3d accelerometer (breathing, activity)
  - Will be used to compute sleep stages and sleep patterns
- Example 2: MoveMonitor – parameters of interest
  - Active Time
  - Total energy expenditure
  - Lying during daytime
  - Movement intensity during walking
  - Physical activity level
  - Steps



128

Personal Informatics Example



Adapted from a framework developed by the Markle Foundation

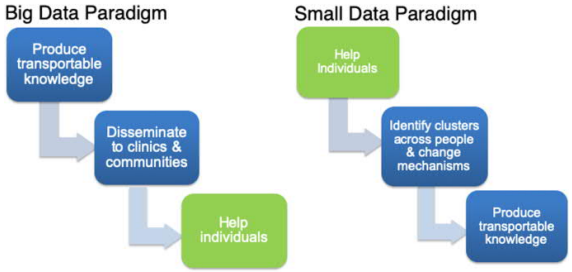
This Photo by Vince Kuraitis is licensed under CC BY-SA

Personal Data as a Boundary Object

- (personal) data can be seen as a *boundary object*
  (Star and Griesemer 1989; Star 2010)
  - many ways in which different communities refer to and think of it
  - E.g. *data trails* as *small data* (Estrin 2013) with *"N = me"*
  - *participatory data* (Shilton 2012) in health
  - *microdata* (Kum et al. 2014) in population informatics
  - *digital footprint* (Madden et al. 2007) in the digital economy

https://www.interaction-design.org/literature/book/the-encyclopedia-of-human-computer-interaction-2nd-ed/human-data-interaction

Big Data & Small Data



- Individual-relevant vs. Aggregate-relevant
- Hypothesis-driven vs. Data-driven

This Photo CC BY

Hekler, E. B., Klasnja, P., Chevance, G., Golaszewski, N. M., Lewis, D., & Sim, I. (2019). Why we need a small data paradigm. *BMC Medicine*, *17*(1), 133. https://doi.org/10.1186/s12916-019-1366-x

Human-Data Interaction (HDI)

- Elmqvist (2011): the *"human manipulation, analysis, and sensemaking of large, unstructured, and complex datasets."*
- HDI is about federating disparate personal data sources and enabling user control over the use of "my data" (McAuley, Mortier, and Goulding 2011)
- *Relatively young area, so terminology and definitions still emerging / changing*

Elmqvist, N., 2011. Embodied Human-Data Interaction. ACM CHI 2011 Work. "Embodied Interact. Theory Pract. HCI 104–107.

Human-Data Interaction (HDI)

Mortier et al. (2014): focus on personal and open data:

- Legibility: process of understanding and making data and analytics algorithms both transparent and comprehensible to people
- Agency: power of handling data; capacity to control, inform, and correct data and inferences
- Negotiability: dynamic relationships that emerge regarding data; understanding and attitudes change over time; reevaluate

Mortier, R., Haddadi, H., Henderson, T., McAuley, D., Crowcroft, J., 2014. Human-Data Interaction: The Human Face of the Data-Driven Society. SSRN Electron. J. https://doi.org/10.2139/ssrn.2508051

HDI Research Topics & Challenges

- Personal data, data ownership and consent
- Embodied interactions
- Data visualization, mining, and analytics *[actually 3 topics]*
- HDI for specific domains (health informatics, smart cities, ..)

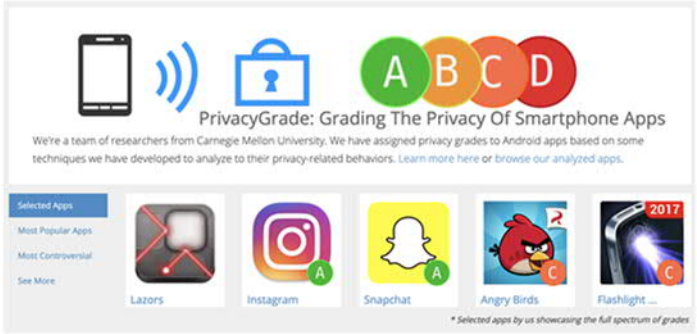
Challenges (selected):

- Users' engagement through participation in design processes
- Models and value of data ownership, social / cultural aspects
- Transcending human and machine limitations in data analysis
- Data influence in decision-making process
- Systemic view of the complete data life cycle

Victorelli, E. Z., Dos Reis, J. C., Hornung, H., & Prado, A. B. (2020). Understanding human-data interaction: Literature review and recommendations for design. *International Journal of Human-Computer Studies, 134*, 13–32. https://doi.org/10.1016/j.ijhcs.2019.09.004

Privacy Expectations

vary by culture, country, application area, type, etc.



Unconsented Data Transfusions

- Attitudes Towards Extracting Personal Device Data for Public Health Emergencies
- Foreshadows embodiment...

Watson, C., & Smeddinck, J. D. (2020). Unconsented Data Transfusions: Attitudes Towards Extracting Personal Device Data for Public Health Emergencies. *Proceedings of Mensch Und Computer 2020*.

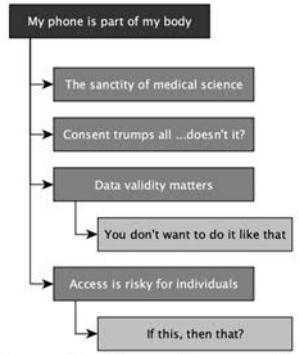


Figure 2: Overarching theme, themes and subthemes.

Key Considerations

- Transparency and audit
  - What audit trails and information are to be provided to support this?
- Privacy and control
  - How can the resulting audit data be used to enable interaction around control of access to and processing of data?
- Analytics and commerce
  - How can the analysis algorithms that are used be made transparent to users (often while retaining protected commercial status)?
- Data to knowledge
  - How can the vast amount of data be used to benefit the individuals and let the society exploit the wealth of information offered by shared data?

via https://haiicmu.github.io

137

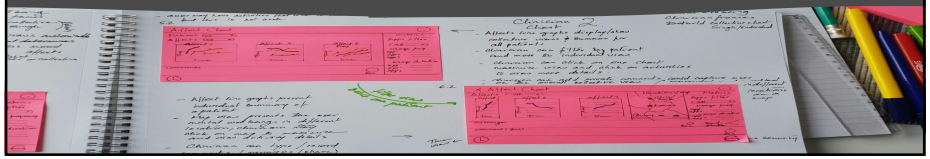
137

WellnessTalk: Concepts

Open Lab

Collaboration with Peter Gallagher, Raghda Marai Saleh Zahran & Jay Rainey...

- personal data is subject to privacy and users' consent
- behavior influences affective states (feelings) and mood
  - daily and online activities influence mood and wellbeing
- mental health and wellbeing is influence by their mood
- affective states are interconnected
- activities include social media threads, health activities, etc.



138

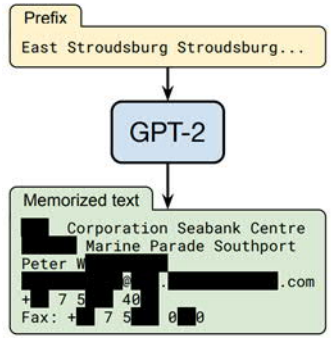
WT

Data-driven, individualised mental health and wellbeing

139

Sensitive Data (= Almost All Data!)

- Fingerprint character?
- All data is health-data...
- It can be possible to extract data from learning models ...
  - Veale, M., Binns, R., & Edwards, L. (2018). Algorithms that remember: Model inversion attacks and data protection law. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *376*(2133), 20180083. https://doi.org/10.1098/rsta.2018.0083
  - Carlini, N., Tramer, F., Wallace, E., Jagielski, M., Herbert-Voss, A., Lee, K., Roberts, A., Brown, T., Song, D., Erlingsson, U., Oprea, A., & Raffel, C. (2020). Extracting Training Data from Large Language Models. *ArXiv:2012.07805 [Cs]*. http://arxiv.org/abs/2012.07805
- ... consider "differential privacy"



140

140

(Health)Data Privacy

The Paradox of Privacy

The More We Reveal, the More Privacy We Desire

https://www.interaction-design.org/literature/book/the-encyclopedia-of-human-computer-interaction-2nd-ed/human-data-interaction

141

This Photo by Dan Utter is licensed under CC BY-SA-NC

141

(Health) Data Interaction

Open Lab

From ongoing work by Jack Holt (supervisee; co-supervisor James Nicholson):

- Growing awareness of the importance of good security practices
- Little consideration of digital legacy after we die
- WS study: Post-mortem privacy paradox
  - users understand importance of planning yet find it uncomfortable
- tension between use of security tools during life and facilitating post-mortem access to chosen assets

→ *facilitate and encourage digital legacy planning while promoting good security habits during life*

J.Holt3@newcastle.ac.uk

142

142

Post-Mortem Privacy Paradox

- work with users of password managers to explore views on the sharing, security and privacy of common digital assets
- when facing loss of control (e.g. care towards end of life and death)
- users recognise value in planning for their digital legacy
- yet: avoid actively doing so
- tension between the use of recommended security tools during life and facilitating appropriate post-mortem access to chosen assets

Jack Holt, James Nicholson, & Jan Smeddinck. (2021, *accepted*). From personal data to digital legacy: Exploring conflicts in the sharing, security and privacy of post-mortem data. WWW '21: Proceedings of The Web Conference 2021. WWW '21: The Web Conference 2021, Ljubljana, Slovenia

143

143

Data Rights Initiatives & Law (e.g. GDPR)

- rules relating to the protection of natural persons with regard to the processing of personal data and rules relating to the free movement of personal data
- protects fundamental rights and freedoms of natural persons and in particular their right to the protection of personal data
- E.g. reason to store / process data, consent, simple terms, clear processing guidelines, consider time requirements, data subjects / controllers / processors, EU (or equivalency) vs. non-EU, …

144

144

Data Rights Initiatives & Law (e.g. GDPR)

- Right to be forgotten
  - … right to have data records pertaining to individuals removed (e.g. from search engines)
  - … but what about impact on group models / derived data?
- Transparency of use → some level of "right to explainability" (more on that in week 3)
- Generally also legal protections "against biased models" based on discrimination protection
- May appear "hindering" from engineering POV, but generally helpful towards explainable and fair AI …

145

145

(Health) Data Interaction

Open Lab

From ongoing work by Alex Bowyer (supervisee):

- Data is used in care, but it disempowers
- Ppl already disempowered, staff/gov/others want more data & more linkage
- Focus on staff needs, staff efficiency
  - not looking at relationships
- HDI, agency
  - data consolidation will make it worse
  - need shared HDI

→ Digital Civics, shared stakeholder needs, find holistic way forward …. also GDPR data explorer

A.Bowyer2@newcastle.ac.uk

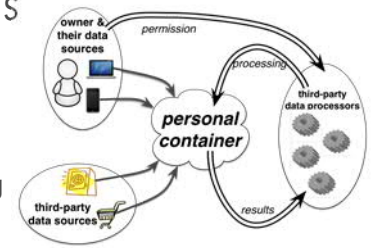
See session video on Personal Data Use

146

146

Data Silos → Data Pods

- Or "personal data vaults" (PDVs)
- E.g. https://solidproject.org/ … https://www.mydata.org/
- Empower individuals by improving their right to self-determination regarding their personal data…
- If "data is the new oil", then arguably, this requires rethinking asset ownership



Author/Copyright holder: Richard Mortier. Copyright terms and licence: CC BY-NC-ND

Mun, M., Hao, S., Mishra, N., Shilton, K., Burke, J., Estrin, D., Hansen, M., & Govindan, R. (2010). Personal data vaults: A locus of control for personal data streams. *Proceedings of the 6th International COnference*, 1–12. https://doi.org/10.1145/1921168.1921191

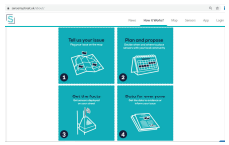
147

147

Democratising Data Science

- Focus on simplicity / reducing effort
- Self-determined human actors (link SDT, Deci, 1975)
- Not only in applications development but also in analytics
  - "democratising data science"
  - data-in.place (Aare Puussaar)

Puussaar, A., Johnson, I. G., Montague, K., James, P., & Wright, P. (2018). Making Open Data Work for Civic Advocacy. *Proceedings of the ACM on Human-Computer Interaction*, *2*(CSCW), 143:1-143:20. https://doi.org/10.1145/3274412



148

148

Human-Data Interaction Summary

An emerging (sub-)field: HDI is about …

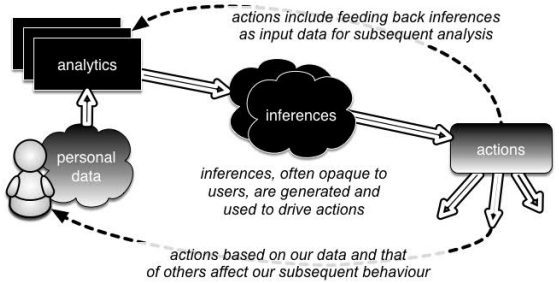
- federating disparate personal data sources and enabling user control over the use of "my data" (McAuley, Mortier & Goulding 2011)
- human manipulation, analysis, and sense-making of large, unstructured, and complex datasets (Elmqvist 2011)
- processes of collaboration with data and the development of communication tools that enable interaction (Kee et al. 2012)
- providing access and understandings of data that is about individuals and how it affects them (Mashhadi, Kawsar & Acer 2014)

https://www.interaction-design.org/literature/book/the-encyclopedia-of-human-computer-interaction-2nd-ed/human-data-interaction

149

149

Human-Data Interaction Summary



we lack **legibility**
means to inspect and reflect on "our" data, to understand its collection and processing

we lack **agency**
means to manage "our" data and access to it, enabling us to act effectively in these systems as we see fit

we lack **negotiability**
means to navigate data's social aspects in collaboration with others and their policies

Author/Copyright holder: Richard Mortier. Copyright terms and licence: CC BY-NC-ND

150

150

Human-AI Interaction & Futures

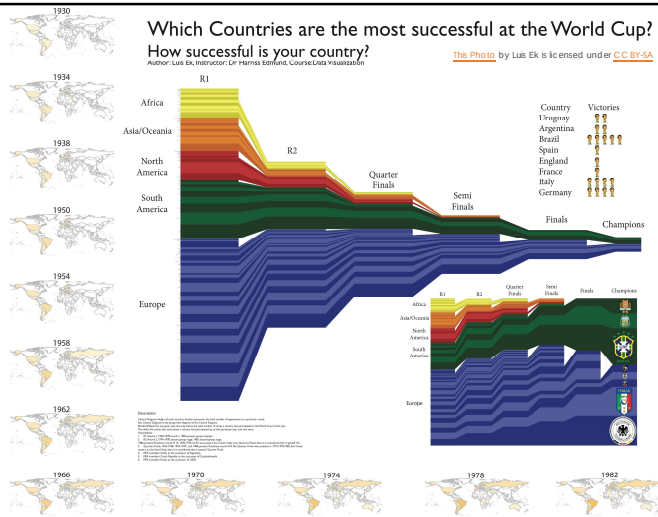
Week 01 – Session 05: Data Visualisation & Communication

Jan Smeddinck

Unless otherwise noted, all materials: (CC BY-NC-SA 4.0)

151

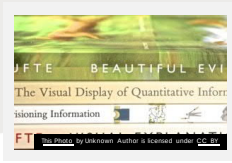
151



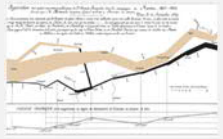
152

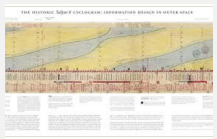
152

Edward Tufte



This Photo by Unknown Author is licensed under CC BY





- *American statistician, pioneer of data visualisation…*

153

This Photo by Unknown Author is licensed under CC BY-NC-ND

153

Why consider visualization in HAII?

- Visualisations may eliminate / reduce need for AI
  - Information visualisation
  - Augmented sensemaking
  - Akin to good interaction design, which can reduce manual effort
- Can visualizing models (and explore data when building)
  - "Understanding" models for developers
  - "Understanding" models for users
    - E.g. describing, predicting outcomes, etc.
- AI can help with creating better visualizations
  - with lots of complex data, difficult to do manually
  - E.g. decision trees

154

154

Visualisation Goals

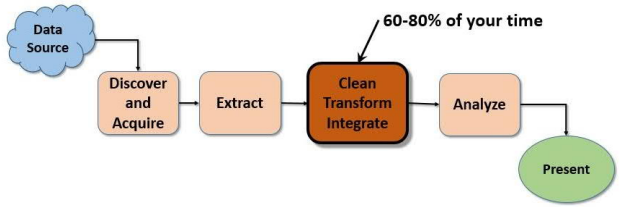
- Exploration
- Sense-making
- Communication
- Aesthetic / experience?
- …

155

155

Visualisation Process / Pipeline

Choose Data → Transform → map to visual attributes



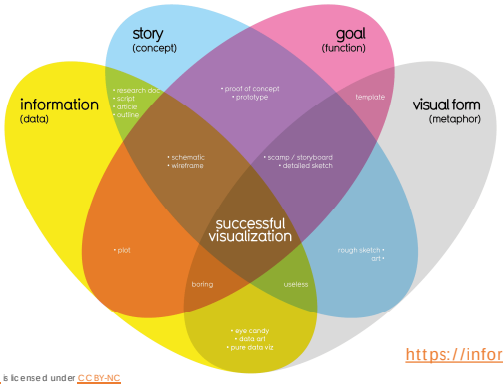
This Photo is licensed under CC BY-NC

Martin Magdinier, RefinePro

156

156

Designing Data Visualisations



https://informationisbeautiful.net/

This Photo is licensed under CC BY-NC

David McCandless
InformationIsBeautiful.net

taken from new book
Knowledge is Beautiful

find out more
bit.ly/KIB_Books

157

157

Basic Visualisations

Consider...

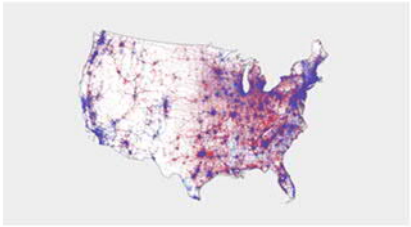
- Spacing / scale
- Light / colour
- Time / dynamics
- Relationships
- Highlighting / guides
- Labels / contextualisation / annotations
- ...

158

158

Picking Appropriate Data

Electoral college map



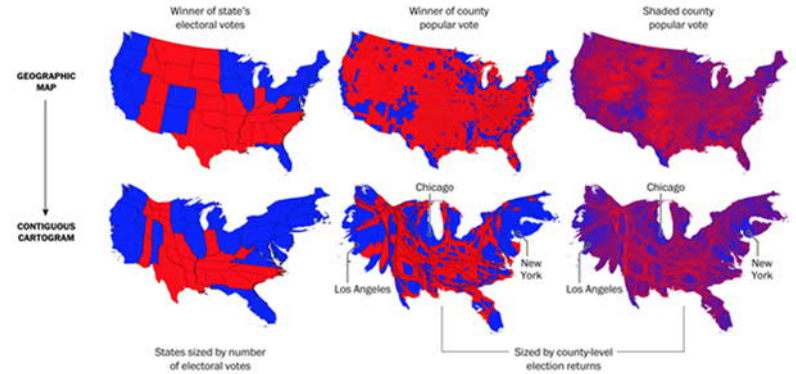
Popular vote map

... for the purpose.

159

159

Picking Appropriate Mapping

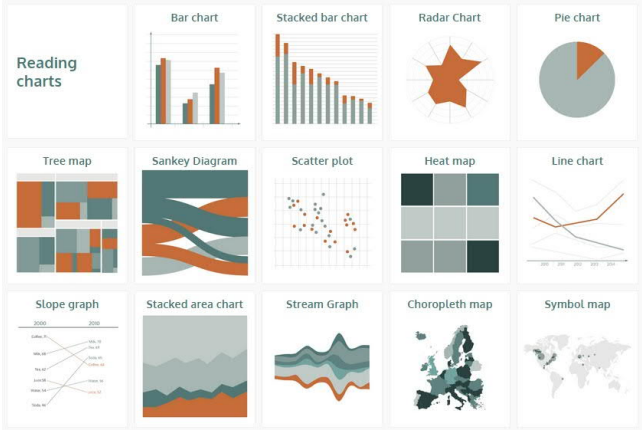


Maps courtesy of Mark Newman

160

160

Data Literacy



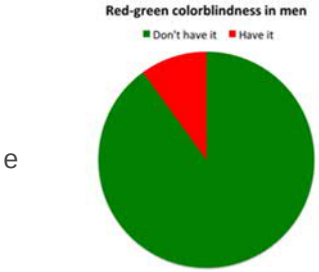
This Photo by Helen Kennedy is licensed under CC BY

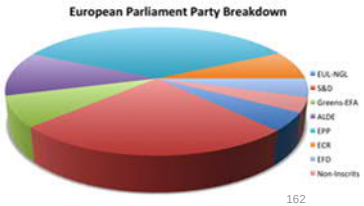
161

161

Representation & Accessibility

- What fraction of men can read the pie chart?
- Reading angles is hard
- Reading angles in perspective is harder
- How much larger is "Don't have it" relative to "have it"?
- Is ALDE > (EPP + Non-inscrits + EUL-NGL)?





162

162

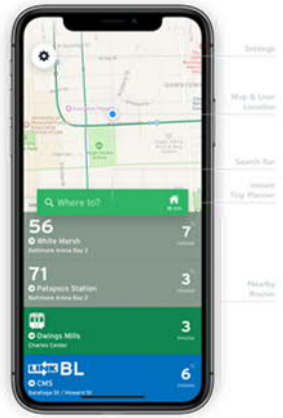
Visualising Uncertainty

- Particularly relevant with regard to communicating uncertainties that may result from learning systems…
- Especially in recommender / decision support system applications

163

163

Worthwhile Visualisations



Where is my bus?

164

164

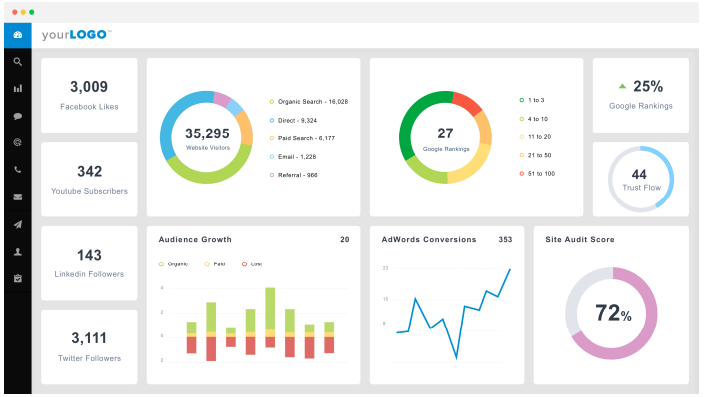
Interactive Visualisations & Dynamics

The best stats you've ever seen | Hans Rosling
https://www.youtube.com/watch?v=hVimVzgtD6w

165

165

Dashboards



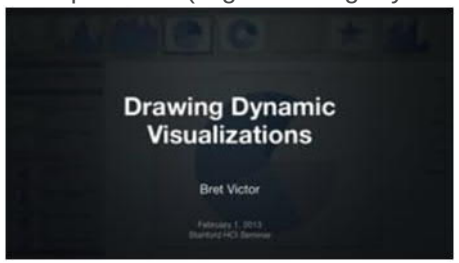
This Photo by Bibhuranjan is licensed under CC BY-SA-NC

166

166

Data Interaction Modalities

- Exploring, Browsing, Searching
- Rich Direct Manipulation (e.g. Drawing Dynamic Vis.)



https://vimeo.com/66085662 and everything else by Bret Victor (https://worrydream.com/)

- *Drowning (in a sea of data)... ;)*

167

167

VR Dataverse (Immersion / Experience)

Open Data Exploration in Virtual Reality x Nordic Tweet Stream (WIP, Nov. 2017) https://vimeo.com/242913692

168

168

Visualisation Empowerment

- Democratising data science → enables civic knowledge and awareness, but also argumentation; this often uses visualisations
- See e.g. work on visualisation empowerment (and VR Data Vis & Ix) by Benjamin Bach (University of Edinburgh)

Further reading:

- Data Visualisation Handbook (Koponen & Hilden) https://datavizhandbook.info/
- Savvy for more?: https://datavis-online.github.io/

169

169

Visualization for developing AI/ML systems

- Play with…
- https://research.google.com/bigpicture/attacking-discrimination-in-ml/
- What-if tool: https://pair-code.github.io/what-if-tool/
- What do you think of these visualizations / tools?
- Where you able to gain insights? Why? Why not?

170

170

Human-AI Interaction & Futures

Week 01 – Session 05:
Foundations of
Human-AI Interaction

Jan Smeddinck

Unless otherwise noted,
all materials: (CC BY-NC-SA 4.0)

171

171

Human-AI Interaction

An emerging term…

- Human-AI Interaction (HAI/HAII)
- Human-Centred/Centric AI (HCAI)
- AI Interaction Design (AIxD)
- AI User Experience (AIUX)

… differences in emphasis and application focus …

172

This Photo by Paul E-Baker is licensed under CC BY-NC-ND

172

HCI + AI & AI + HCI Perspectives

- HCI with AI/ML
  - E.g. interacting with a conversational interface
- AI/ML for HCI
  - E.g. adaptive interfaces
- (HCI for AI/ML)
  - E.g. data collection tools with good UX
  - A bit more in week 3 (human computation)

173

173

Put-that-there (1980)

- How could this be helpful (augment)?
- How could this fail?
- Is it efficient?
- Is it a good UX?

Bolt, R. A. (1980). \"Put-that-there\": Voice and gesture at the graphics interface. *Proceedings of the 7th Annual Conference on Computer Graphics and Interactive Techniques*, 262–270. http://doi.acm.org/10.1145/800250.807503

174

174

Apple Knowledge Navigator (1987)

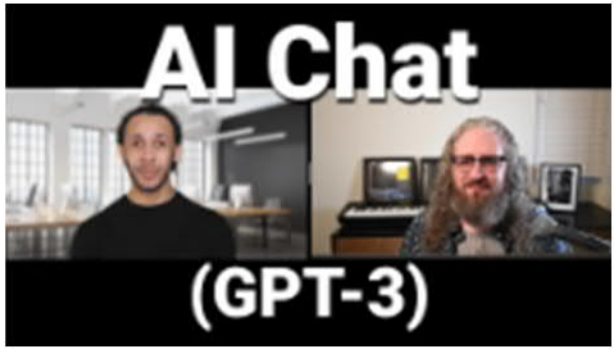
- How could this be helpful (augment)?
- How could this fail?
- Is it efficient?
- Is it a good UX?
- *Is this realistic/achievable?*

https://www.youtube.com/watch?v=umJsITGzXd0

175

175

AI Agents in 2020 (preview)



https://www.youtube.com/watch?v=PqbB07n_uQ4&t=211

176

176

Interacting with Systems that Change

Traditional "tools" don't change by themselves ... if changing:

1. Complex but (more or less) rule-based
2. Learning (pattern-based, may be predictable)
3. Creative / innovating (less predictable)

This Photo by Quentin Monnet is licensed under CC BY

177

177

Personalisation & Customization

Concept	Definition
Adaptability	The fact that a system is not fixed, but can be changed (to the needs of users, to changing environmental contexts, etc.; changes are usually understood to be performed manually).
Customization	The act of changing a system to the needs of a user group or individual user (manually or automatically; may can be done by the group itself or by the user him- or herself, but may also be done by third parties; often related to the appearance or content of the given system).
Personalization	The act of changing a system to the needs of a specific individual user (often automatic but does not have to be, i.e., can be understood as a specific form of customization with a focus on individuality; personalization is also often related to appearance or content).
Adaptivity	The fact that a system is not fixed, but dynamically changes over time (to adjust to the needs of users or an individual user, or to adjust to changing environmental contexts, etc.; typically happens automatically; often related to settings and parameters present in the given system).

Note: can be heuristic, ML based, etc.

Smeddinck, J. (2017). *Human-Computer Interaction with Adaptable & Adaptive Motion-based Games for Health* [University of Bremen]. SUUB Bremen. https://arxiv.org/abs/2012.03309

178

178

Adaptive Systems (e.g. UIs)

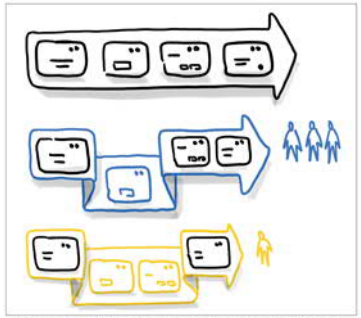


Figure 1. AUIs can learn to automate a frequent task by adapting to groups of people. They can also learn to automate by adapting to the behavior of an individual.

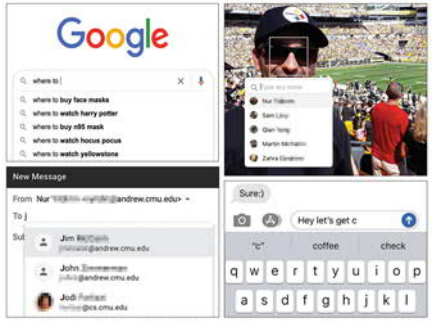


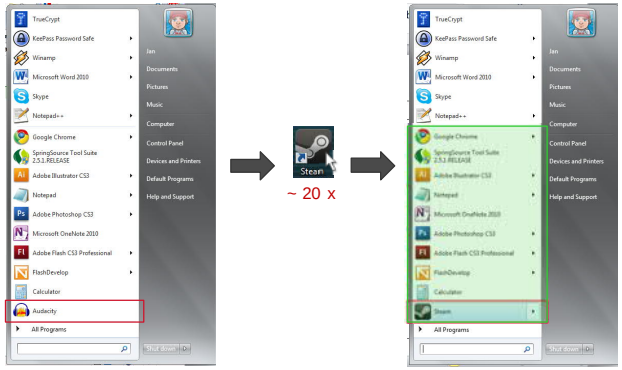
Figure 2. Examples of Ranked List of Likely Targets (clockwise from top left): Google Search's suggestions of likely search terms, Facebook's friend tagger, word suggestion for text messaging, email applications suggestions for message recipients.

Zimmerman, J., Oh, C., Yildirim, N., Kass, A., Tung, T., & Forlizzi, J. (2020). UX designers pushing AI in the enterprise: A case for adaptive UIs. *Interactions*, *28*(1), 72–77. https://doi.org/10.1145/3436954

179

179

Adaptive Interfaces: Where is Audacity!?!

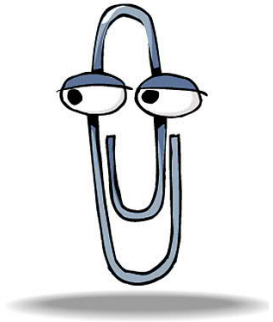


28/12/2021

180

180

Rules → Smartness (!?!)



181

Dix et al's Usability Principles

- Learnability
  - *the ease with which new users can begin effective interaction and achieve best performance*
- Flexibility
  - *the many ways the user and system exchange information*
- Robustness
  - *the level of support provided to the user in determining successful achievement and assessment of goals*

182

Schneiderman's 8 Golden Rules

183

Strive for consistency

Consistent sequences of actions should be required in similar situations; identical terminology should be used in prompts, menus, and help screens; and consistent commands should be employed throughout.

1

184

Nielsen Heuristic Evaluation

1. Visibility of system status
   - keep users informed through feedback
2. Match between system and the real world
   - follow real-world conventions
3. User control and freedom
   - Support undo and redo
4. Consistency and standards
   - follow conventions and standards
5. Error prevention
   - make it difficult for users to make errors

Usually also done by (dedicated) expert!

This Photo by Enrico Signoretti is licensed under CC BY-SA

https://www.interaction-design.org/literature/article/heuristic-evaluation-how-to-conduct-a-heuristic-evaluation

185

Nielsen Heuristic Evaluation

6. Recognition rather than recall
   - Users should not have to remember information, instead instructions should be easily visible or retrievable
7. Flexibility and efficiency of use
   - Allow users to tailor actions so that the design can cater to both experienced and novice users
8. Aesthetic and minimalist design
   - Only relevant information should be displayed
9. Help users recognise, diagnose, and recover from errors
   - error messages should be in plain English and should suggest a solution
10. Help and documentation
    - should be easy to understand and use

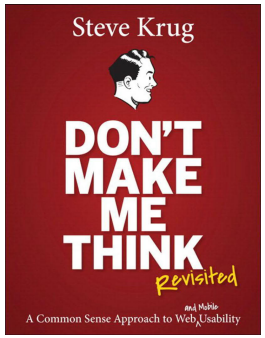
https://www.interaction-design.org/literature/article/heuristic-evaluation-how-to-conduct-a-heuristic-evaluation

186

Remembering just one thing…

- Krug's first law of usability…
- Make things…
  - Self-evident, obvious, self-explanatory
- Think "next door neighbours"…
- The competition / next distraction is always just one click away

Via Krug, 2013



28/12/2021 187

187

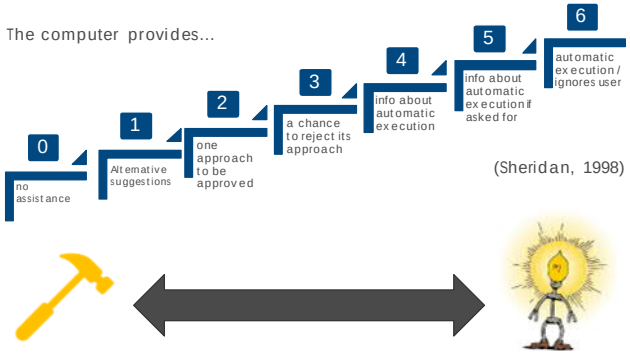
Trunk Test

- Step 1: Choose a page anywhere in the site / app at random, and print it.
- Step 2: Hold it at arm's length or squint so you can't really study it closely.
- Step 3: As quickly as possible, try to find and circle each of these items:
  - Site ID
  - Page name
  - Sections (Primary navigation)
  - Local navigation
  - "You are here" indicator(s)
  - Search

28/12/2021 188

188

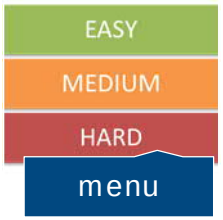
Sheridan's Scale of Automation

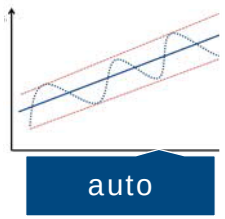


This Photo by Quentin Monnet is licensed under CC BY

189

Difficulty Choices & Player Experience





- Can impact autonomy needs satisfaction
- Other game experience measures (e.g. presence / immersion) appear untouched

Smeddinck, J. D., Mandryk, R. L., Birk, M. V., Gerling, K. M., Barsilowski, D., & Malaka, R. (2016). How to Present Game Difficulty Choices?: Exploring the Impact on Player Experience. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 5595–5607. https://doi.org/10.1145/2858036.2858574

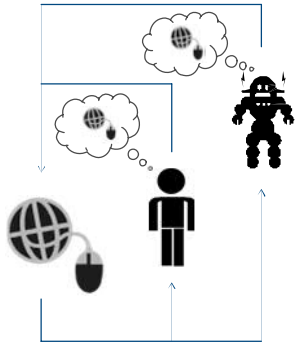
190

Automation vs. Control

191

Learning from Interactions for Adaptive Systems

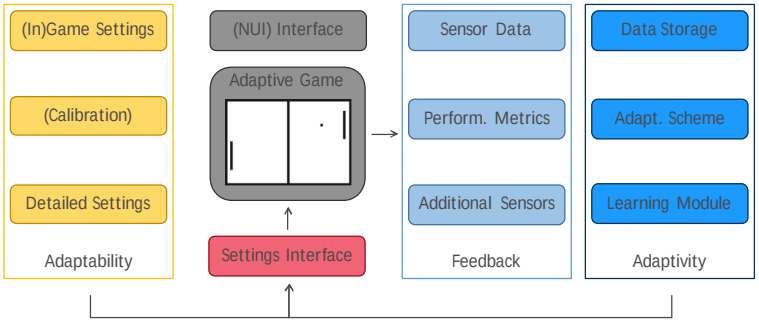
Heuristic adaptive systems usually „reactive" ... to more effectively prevent going in „wrong directions" need to be „predictive" → AI/ML



- Build knowledge about the user, the interface, the agent and the world in terms of properties and capabilities...
- Learn to interpret the user behavior...
- Predict their needs and desires...
- Reason on the outcomes of different potential actions...
- *Aka: contextual computing with context models and feedback...*

192

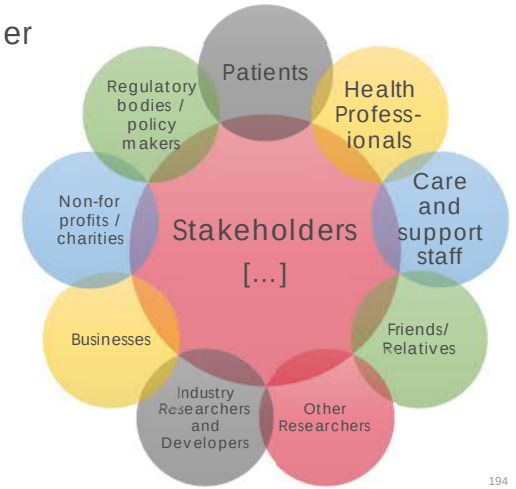
Complex Building Blocks (e.g. Games)



193

193

Complex Stakeholder Landscapes in modern IxD (e.g. Digital Health)



194

194

Complex Context

- The world
- The current situation, social settings, etc.
- The history
- A complex problem ...
- ... dynamic!

195

195

UB/UX vs Uncertainty & Unpredictability

- UB/UX "laws" vs. "AI" systems

Links to direct manipulation...

- Continuous Representation of Object of Interest (feedback)
- Rapid, incremental, and reversible interactions
- Physical Actions

196

196

Potential & Risks

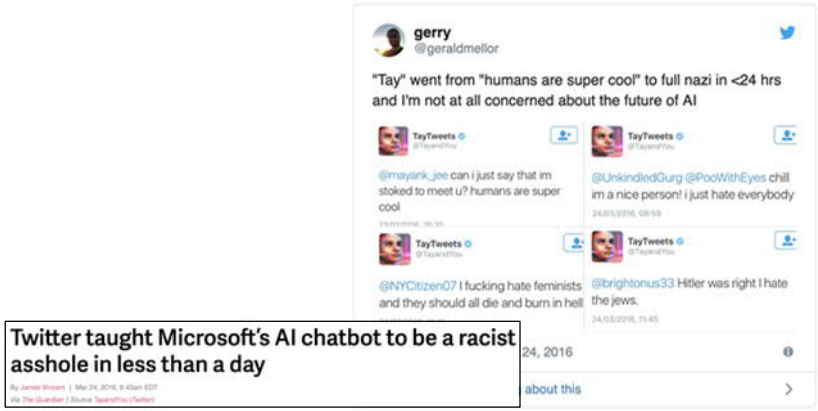
- Relinquishing control to an AI/ML agent can be helpful, but can be much harder to correct or understand if things go wrong
- "Unpredictability" can be a positive thing in one kind of experience, and a terrible thing in another...

https://www.wired.com/story/when-it-comes-to-gorillas-google-photos-remains-blind

197

197

Severe Failure (Tay.ai)



198

What was the error?: *Severe* Failure

- Tay's earlier version XiaoIce ran on China's most widespread instant messaging app Wechat ... without any major ethical incidents
- What makes Twitter a different environment?
- Tay had no *moral agency*. To her, words like Hitler or Holocaust are not different from words like chair or Oklahoma
- 2018 version used black-listing and "moral judgments"... better? Cencorship? Nuance?
- *Same still true for very large language models (ala GPT-3)?*

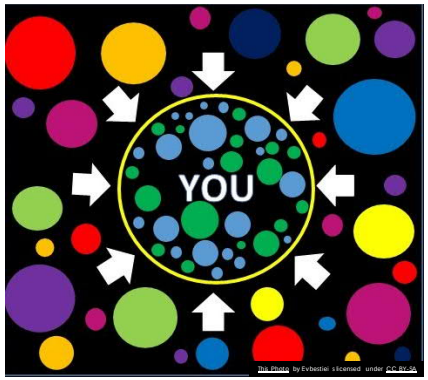
https://chatbotslife.com/the-accountability-of-ai-case-study-microsofts-tay-experiment-ad577015181f

199

199

Memes, Sharing, Liking
Attention Economy
Personalization
Recommender Systems

Filter Bubbles



200

200

Other Errors: Poor model performance

- Usually solvable by acquiring more training data for the situations the model is weakest at
- Data is expensive to collect and prepare, and your company or organization has limited resources. Prioritizing *what* specific data to collect is essential
- Designers can use rule or non-ML based fallbacks to still deliver the user *some* value when model performance isn't good enough for some cases
  - Mixed models (Wizard-of-Oz) often used in business world … in particular start-ups…

201

201

Other Errors: Low confidence or *false* High confidence in a prediction

- Low confidence predictions can mean that the model has lower performance, or the phenomena itself is just… less predictable
- Communicating with the user or providing good non-AI/ML fallbacks is key
  - Think levels of control (can be dynamic)
- High confidence (when the model is really wrong) is worse
  - Unkown unkown errors
- Need to give the user some error correction or feedback method to deal when this happens

202

202

Other Errors: Relevance/appropriateness errors

- Airbnb suggesting 'fun local activities' when you're traveling for a funeral
- Exercise app suggesting 'time to get up and walk!' when you're seated on a long flight
- Amazon suggesting products that you are allergic to or can't eat

203

203

Respectful Diet Information and Decision Support

From ongoing work with Remco Benthem de Grave (supervisee):

- Current diets are unsustainable
  - environmental + health perspective
- Personalized support with digital tech.
- Available systems often coercive (overly persuasive)

→ *Focus: supporting (non-coercively) existing diet values and intentions through personalisation and minimizing effort of making sustainable choices*

r.benthemdegrave2@newcastle.ac.uk

204

204

Different Users & Similar Inputs

1. Use Spotify to play 1970s pop jams at your Mum's party
2. Use Spotify to play your favorite study jams
3. Use Spotify to hate-listen to Ariana Grande (sorry) with your roommate
4. Your roommate also controls the same Spotify account to play their favorite study jams

What music should Spotify recommend this account play?

205

205

Silent Errors? UX?

- Not all errors are visible to the computer
- How do you measure when your device can't hear "OK, Google"
- How does your fitness tracker know it "should have been recording"?
- How does the system know when a user liked/did not like an automated decision?
- *Hypothesis: this is why humans have dialogue … as in Habermas: "Theory of Communicative Action" … but this is a tough challenge … until then …*

206

206

So when to use AI?

Case-by-case decision, but:

- When automation is actually beneficial
- When direct manipulation does not solve the challenge
- When biases can be controlled

AI probably better	AI probably **not** better
❑ The core experience requires recommending different content to different users.	❑ The most valuable part of the core experience is its predictability regardless of context or additional user input.
❑ The core experience requires prediction of future events.	❑ The cost of errors is very high and outweighs the benefits of a small increase in success rate.
❑ Personalization will improve the user experience.	❑ Users, customers, or developers need to understand exactly everything that happens in the code.
❑ User experience requires natural language interactions.	❑ Speed of development and getting to market first is more important than anything else, including the value using AI would provide.
❑ Need to recognize a general class of things that is too large to articulate every case.	❑ People explicitly tell you they don't want a task automated or augmented.
❑ Need to detect low occurrence events that are constantly evolving.	
❑ An agent or bot experience for a particular domain.	
❑ The user experience doesn't rely on predictability.	

https://pair.withgoogle.com/worksheet/user-needs.pdf

207

207

Need to Mitigate Risks & Work With Them

Will discuss…

- Conceptual Foundations
- Then Current Human-AI Interaction Frameworks

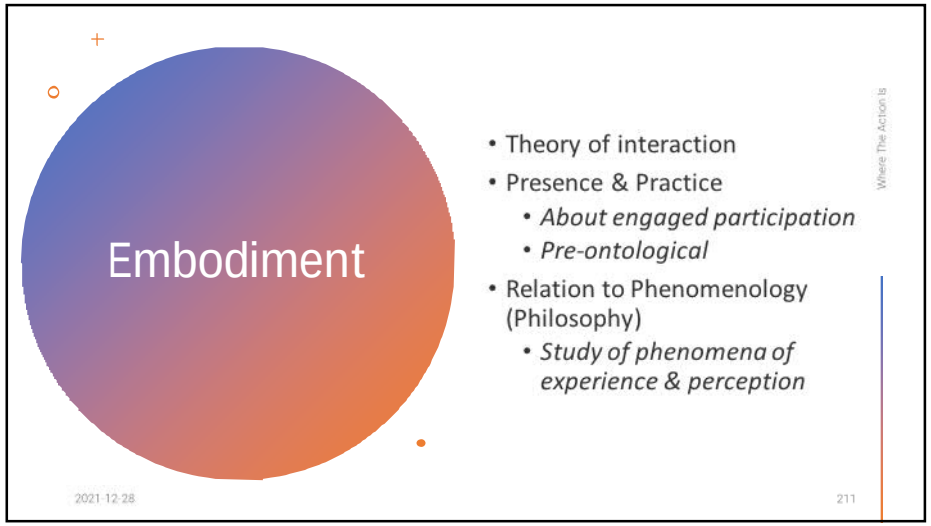
Starting out where conceptual foundations of HCI/UB/UX/IxD typically trail off (after basics of perception, interaction, etc.)

208

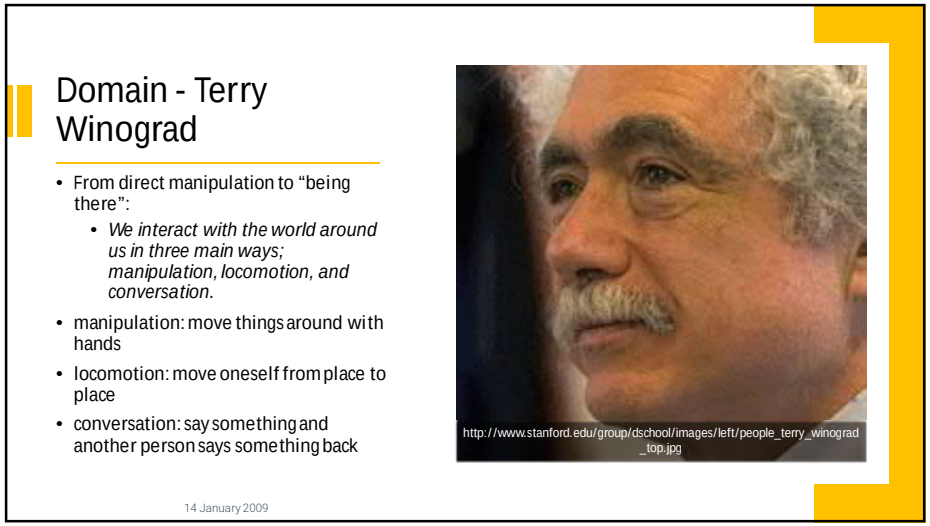
208

209

210



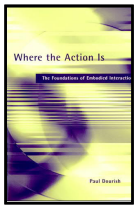
211



212

Embodiment: Into HCI & IxD

- Where The Action Is:
  The Foundations of Embodied Interaction
  - 2001, Cambridge, MIT Press
- Paul Dourish:
  - Prof. of Informatics @ Univ. of California
  - Glasgow > Uni. Edinburgh > Cambridge (EuroPARC) > London (UCL, PhD) > Palo Alto (Xerox) > Cupertino (Apple) > Los Angeles (UCI)
    (http://www.dourish.com/)



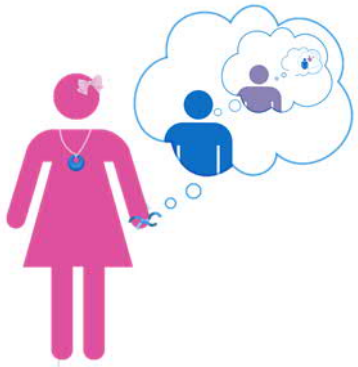
http://www.flickr.com/photos/pdourish/3103575261/sizes/o/
http://mitpress.mit.edu/images/products/books/0262041960.jpg

Embodiment: Phenomenology

Away from Cartesian dualism…

Edmund Husserl	Martin Heidegger	Alfred Schutz	Maurice Merleau-Ponty	Ludwig Wittgenstein
• Founder of Phenomenology • From abstract Galilean science to things that matter • Questions of experience, memory, mind, cognition	• Ending Cartesianism (separation of inner mental life and outside world) • Dasein (being-in-the-world) • No theory prior to praxis	• The Phenomenology of the Social World • Life-world & intersubjectivity	• The Phenomenology of Perception • Body + phisical & social skills • Embodiment	• Semiotician • „The meaning of a word is how we use it." • There is no truth • language games

Embodiment: Theory of Mind



http://blogcraftzine.com/BrainBunny.jpg

Embodiment: Mirror Neurons

https://sites.psu.edu/psych256su16-2/2016/06/28/mirror-neurons/

Affordances

- *"what the environment offers the individual"*
- Coined by James J. Gibson (1966)
- Good design <-> clear affordances
- Read more: Don Norman (2013)

Tea set image by: https://commons.wikimedia.org/wiki/File:2004_MujiWanoSyokki-TeasetMug_Masahiro-Mori.jpg

Affordances

- Chairs?
- Light switches?
- Smart cards?

Door handle image by: https://www.flickr.com/photos/cogdog/7313104802

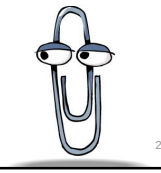
Anthropomorphologisation

This Photo by TV Tropes is licensed under CC BY-SA

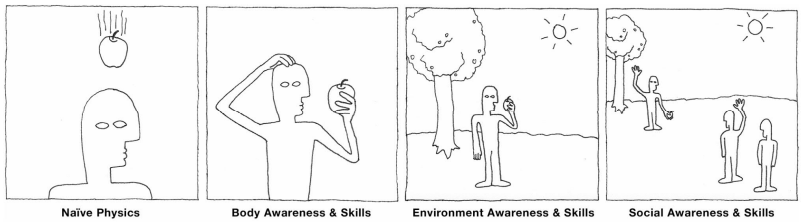
This Photo by Unknown Author is licensed under CC BY-SA-NC

This Photo by smilekiddo is licensed under CC BY-NC-ND

- "The computer did not like me today."
- "It decided to crash..."
- "I feel betrayed by my word editor."
- "It does not want to do what I want it to do."
- ...



Reality-based interaction (RBI)



Jacob, R. J. K., Girouard, A., Hirshfield, L. M., Horn, M. S., Shaer, O., Solovey, E. T., & Zigelbaum, J. (2008). Reality-based interaction: A framework for post-WIMP interfaces. *Proceedings of the Twenty-Sixth Annual SIGCHI Conference on Human Factors in Computing Systems*, 201–210.

Social Interaction

- Social cognition
- hive mind / wisdom of the crowd
- memes

This Photo by Earlham College is licensed under CC BY-NC

Bubble: https://commons.wikimedia.org/wiki/File:Crowdtesting.jpg

221

221

Social Context

- other people
  - status (bosses don't type)
  - showing off
  - competition
  - fear of failure
- motivation
  - fear
  - allegiance
  - ambition
  - self satisfaction
- inadequate systems
  - frustration
  - lack of positive motivation

222

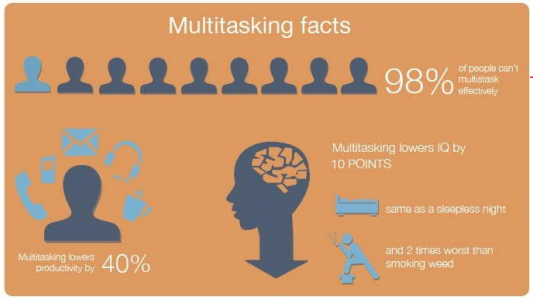
222

Collaboration

- A "reason" for us being social
- Shared intentions
- Joint-action (deeply engrained)
- CSCW
- Key concepts: trust, reliability, responsibility
- *We apply principles to non-human sentient beings (e.g. history of human <-> dog collaboration)*

223

223



Cog. Dims.: Multitasking

- Modern computers can be really good at it
- XX chromosome carriers are worse
- XY chromosome carriers are even worse!
- Everyone is actually "somewhere in between" and despite population averages individuals can be anything...

224

224

Cognitive Dimensions / Aspects

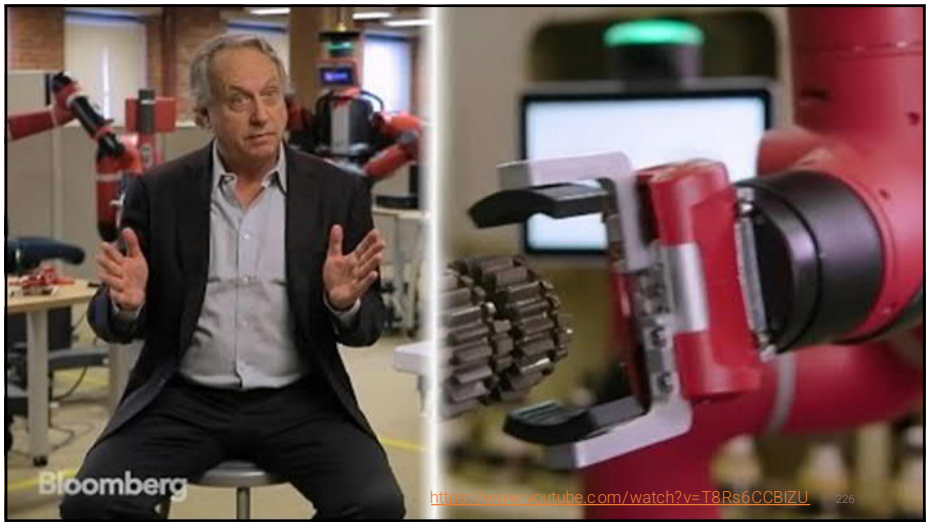
- learnings from what we know about cognition applied to design considerations...

http://blog.craftzine.com/BrainBunny.jpg

http://blog.craftzine.com/BrainBunny.jpg

https://sites.psu.edu/psych256su16-2/2016/06/28/mirror-neurons/

225



226

Embodiment: Drawn Together

- Focus on human skills & activities
- Participation in the world (physical and social reality)
- Spread in space & time (not system time)

→ Context matters! (actions <-> settings)

Also consider: Gallagher:
„How the body shapes the mind“

Charles Sanders Peirce, semiotician, look him up about the importance of context for meaning!

2021-12-28 227

227

Setting the Scene

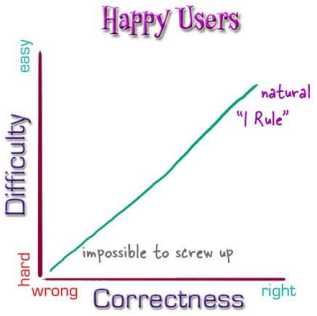
- How is the system presented relative to its capabilities?
- How does user control work, how is it communicated?
- For complex interactions: are there repair strategies?
  - Undo / “go back” etc.?
- `Expression of uncertainties?
- Living with imperfections, recovering from errors?

... can use the concepts from above to inform design thinking ...

228

228

Most fundamentals from interaction design still apply … but not all heuristics / "rules" do …



"Make the right things easy and the wrong things hard"

http://headrush.typepad.com/creating_passionate_users/2005/10/making_happy_us.html

229

Social Factors in which AI can Excel…

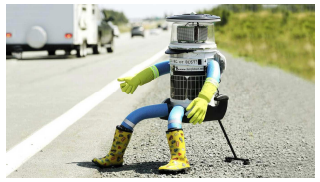
- Motivation (e.g. self-determination theory)
  - Competence, autonomy, relatedness
- Interesting, versatile, exciting
  - Random reward schemes (see "operant conditioning" after BF Skinner with variable ratio reinforcement schedule)
    cf. e.g. https://psyc1100.pressbooks.com/chapter/operant-conditioning/
- Collaboration, augmentation, integration …
- *… examples of designing with social factors in mind?*

230

230

HelpMe Robot Journeys

- hitchBOT (2013 – 2015)
- *Can robots trust humans?*
- Gained international attention for successfully hitchhiking across Canada, Germany and the Netherlands … ☺
- … but in 2015 its attempt to hitchhike across the United States ended prematurely when the robot was stripped and decapitated in Philadelphia, Pennsylvania ☹
- Now actively used in delivery robots!



This Photo by the digeon is licensed under CC BY-NC-ND

This Photo by the digeon is licensed under CC BY-NC-ND

Wood, L. J., Zaraki, A., Robins, B., & Dautenhahn, K. (2019). Developing Kaspar: A Humanoid Robot for Children with Autism. *International Journal of Social Robotics*. https://doi.org/10.1007/s12369-019-00563-6

231

231

Otherware

- People perceive autonomous systems / agents as *counterparts*
- From *embodied* relationship to *alterity*
- Technology becomes other…
- We call this class of interactive systems *otherware* (https://otherware.net)

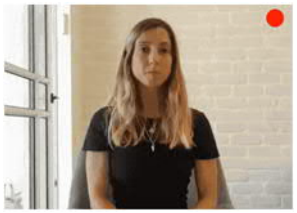
Hassenzahl, M., Borchers, J., Boll, S., Pütten, A. R. der, & Wulf, V. (2021). Otherware: How to best interact with autonomous systems. *Interactions*, *28*(1), 54–57. https://doi.org/10.1145/3436942

232

232

HAll Incredible Opportunities: Accessibility

- Dual-relationship with AI (as noted in many other areas):
  - Making AI systems accessible
  - But also: AI systems that can make other things accessible
- E.g. "Lookout" Android app to help visually impaired identifying packaged foods
  - powerful-but-compact machine learning model
  - real-time on a phone
  - ~ 2 million products
- E.g. automatic ASL detection for video conf.



Moryossef, A., Tsochantaridis, I., Aharoni, R. Y., Ebling, S., & Narayanan, S. (2020). *Real-Time Sign Language Detection using Human Pose Estimation.*

https://ai.googleblog.com/2020/07/on-device-supermarket-product.html

General AI/ML Model Construction

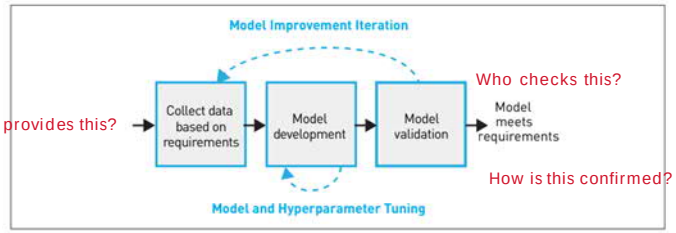


Figure 2. The machine learning train-validation-test cycle.

Le, H. V., Mayer, S., & Henze, N. (2021). Deep learning for human-computer interaction. *Interactions*, *28*(1), 78–82. https://doi.org/10.1145/3436958

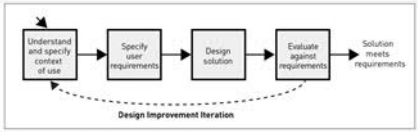


Figure 1. The UCD process as described in ISO 9241-210 [1].

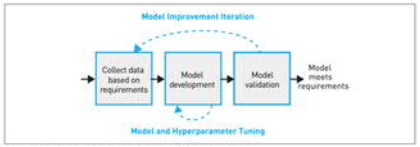


Figure 2. The machine learning train-validation-test cycle.

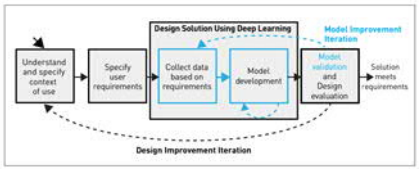


Figure 3. Adapted UCD process for deep learning in the context of interactive systems in HCI.

UCDDL

- User centred design process for deep learning
- Core idea:
  - Waterfall → iterative
  - And user-centred
- Work in progress:
  - Participatory and co-design with learning systems?

Le, H. V., Mayer, S., & Henze, N. (2021). Deep learning for human-computer interaction. *Interactions*, *28*(1), 78–82. https://doi.org/10.1145/3436958

Active area of research...

Horvitz, Eric. "Principles of mixed-initiative user interfaces." In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, pp. 159-166. ACM, 1999. Pdf

Saleema Amershi, Dan Weld, Mihaela Vorvoreanu, Adam Fourney, Besmira Nushi, Penny Collisson, Jina Suh et al. "Guidelines for human-AI interaction." In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, p. 3. ACM, 2019.

Rafal Kocielnik, Saleema Amershi, and Paul N. Bennett. "Will You Accept an Imperfect AI?: Exploring Designs for Adjusting End-user Expectations of AI Systems." In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, p. 411. ACM, 2019. Pdf

Gagan Bansal, Besmira Nushi, Ece Kamar, Daniel S. Weld, Walter S. Lasecki, and Eric Horvitz. "Updates in human-ai teams: Understanding and addressing the performance/compatibility tradeoff." In *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, pp. 2429-2437. 2019. Pdf

Gagan Bansal, Besmira Nushi, Ece Kamar, Walter S. Lasecki, Daniel S. Weld, and Eric Horvitz. "Beyond Accuracy: The Role of Mental Models in Human-AI Team Performance." In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, vol. 7, no. 1, pp. 2-11. 2019. Pdf

Daniel S. Weld, and Gagan Bansal. "The challenge of crafting intelligible intelligence." *Communications of the ACM* 62, no. 6 (2019): 70-79. Pdf

Besmira Nushi, Ece Kamar, and Eric Horvitz. "Towards accountable ai: Hybrid human-machine analyses for characterizing system failure." In *Sixth AAAI Conference on Human Computation and Crowdsourcing*. 2018. Pdf

Forough Poursabzi-Sangdeh, Daniel G. Goldstein, Jake M. Hofman, Jennifer Wortman Vaughan, and Hanna Wallach. "Manipulating and measuring model interpretability." *arXiv preprint arXiv:1802.07810* (2018). Pdf

Human-AI Interaction Design Frameworks

Big corp. research arms chiming in…

- Microsoft: https://www.microsoft.com/en-us/research/project/guidelines-for-human-ai-interaction/
- Google: https://pair.withgoogle.com/guidebook/
  - Focus on responsible AI … more later
- IBM: https://developer.ibm.com/technologies/machine-learning/articles/machine-learning-and-bias/
  - Focus on preventing bias …

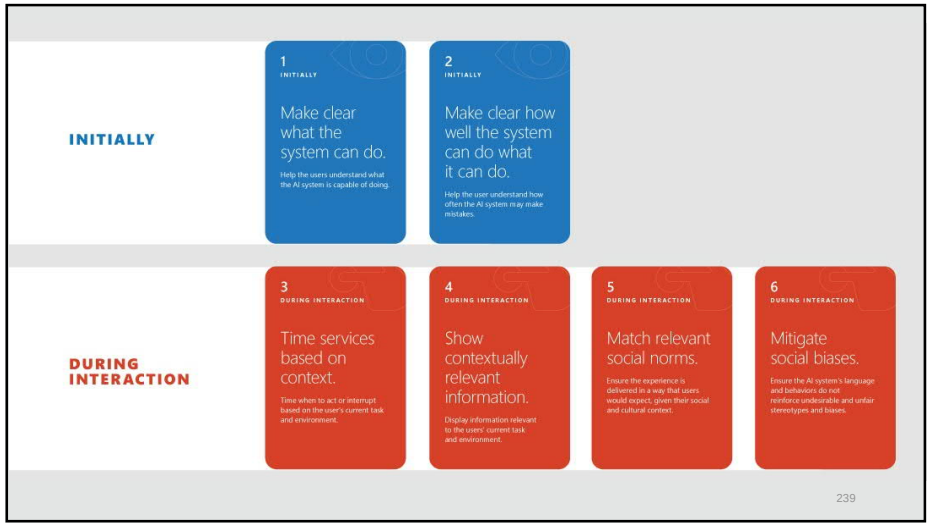
237

237

Guidelines for Human-AI Interaction

238

238



239

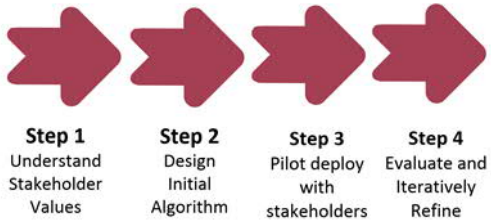
239

240

240

Value sensitive algorithm design...

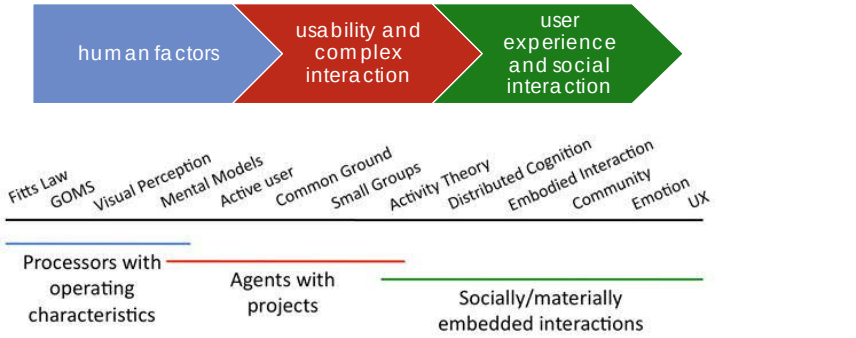
(Zhu, Yu, Halfaker & Terveen, CSCW 2018)



Evaluate: algo. accuracy + stakeholders: acceptance, impact

241

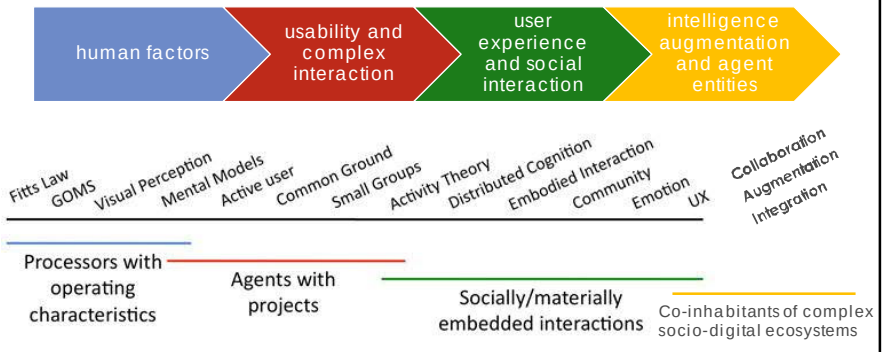
History of HCI



via: https://www.interaction-design.org/literature/book/the-encyclopedia-of-human-computer-interaction-2nd-ed/human-computer-interaction-brief-intro

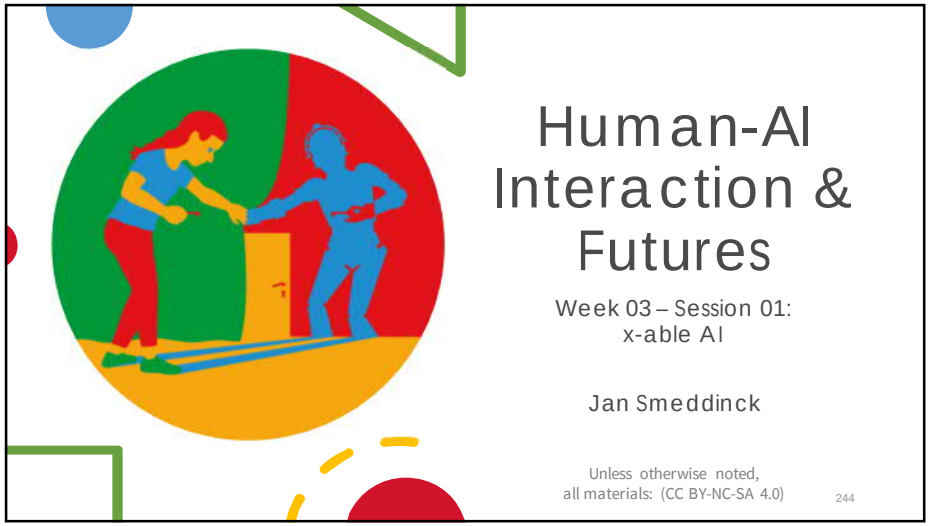
242

History of HCI + AI ???



via: https://www.interaction-design.org/literature/book/the-encyclopedia-of-human-computer-interaction-2nd-ed/human-computer-interaction-brief-intro

243



Human-AI Interaction & Futures

Week 03 – Session 01:
x-able AI

Jan Smeddinck

Unless otherwise noted,
all materials: (CC BY-NC-SA 4.0)

244

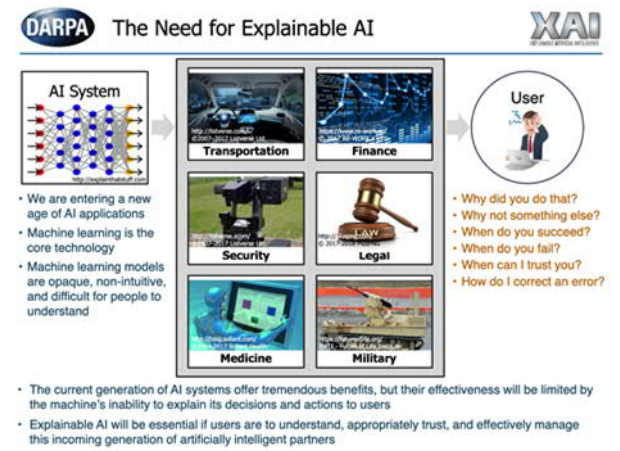
Learning Goals

Brief outline of relevant concepts with pointers to relevant further materials:

- Contextualise motivation / need for explainable AI
- Development of the area and key principles
- Explainable / x-able AI
- Links to concepts from interaction design (e.g. UB/UX)
- Links to challenges in AI

245

245



Via: https://www.darpa.mil/attachments/XAIProgramUpdate.pdf

246

Explainable AI (XAI)

- EU General Data Protection Regulation (GDPR) stipulates right to obtain "meaningful information about the logic involved" - commonly interpreted as a "right to an explanation" - for consumers affected by an automatic decision (Parliament and Council of the European Union, 2016)
- Developing field: no clear agreement about what an explanation is, nor what a good explanation entails
- First emerged in mid-1980s among wave of "expert systems" (remember: move to more practical applications)

Confalonieri, R., Coba, L., Wagner, B., & Besold, T. R. (2021). A historical perspective of explainable Artificial Intelligence. *WIREs Data Mining and Knowledge Discovery*, *11*(1), e1391. https://doi.org/10.1002/widm.1391

247

247

Explainable AI (XAI)

- Different approaches in expert systems / ML / recommenders / neuro-symbolic learning and reasoning

Confalonieri, R., Coba, L., Wagner, B., & Besold, T. R. (2021). A historical perspective of explainable Artificial Intelligence. *WIREs Data Mining and Knowledge Discovery*, *11*(1), e1391. https://doi.org/10.1002/widm.1391

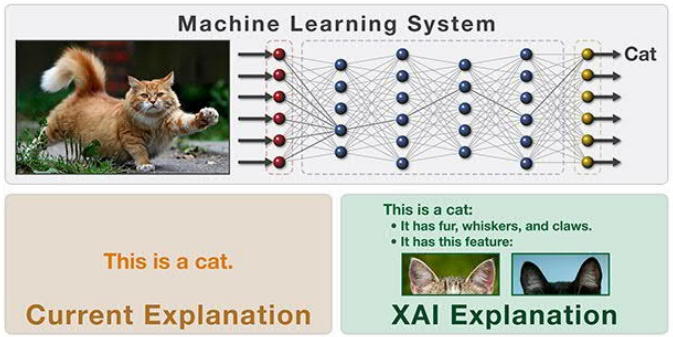
- Social motivation, but also commercial motivation
- *Explainability can be an excuse…*

Applied examples (code):
https://github.com/jphall663/interpretable_machine_learning_with_python

248

248

Explainable AI (XAI)



This Photo by Unknown Author is licensed under CC BY-SA-NC

249

249

Variable Importance Plots

- SHAP Values (SHapley Additive exPlanations)
- Impact of each feature:
  - interpret impact of having a certain value for a given feature in comparison to the prediction if that feature took baseline value

E.g.: predict whether a team will have a player win the Man of the Match award...

See: https://www.kaggle.com/dansbecker/shap-values
Theory brief: https://towardsdatascience.com/one-feature-attribution-method-to-supposedly-rule-them-all-shapley-values-f3e04534983d

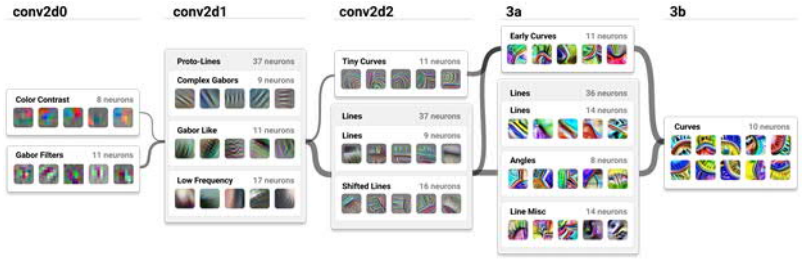
Lundberg, S. M., & Lee, S.-I. (2017). A Unified Approach to Interpreting Model Predictions. *Advances in Neural Information Processing Systems, 30*. https://papers.nips.cc/paper/2017/hash/8a20a8621978632d76c43dfd28b67767-Abstract.html

250

250

Complex Parameter Spaces?

- Progress in building deeper understandings of workings of complex ANNs, but not explainable in the sense of "understandable for the regular consumer":

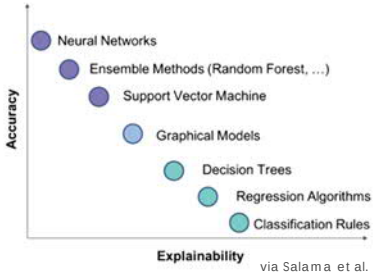


Cammarata, N., Goh, G., Carter, S., Voss, C., Schubert, L., & Olah, C. (2021). Curve Circuits. *Distill*, *6*(1), e00024.006. https://doi.org/10.23915/distill.00024.006s

251

251

Accuracy vs. Explainability



via Salama et al.

- Less performant model may be more accurate when including human operator / use...
- Care about performance of AI + human using it

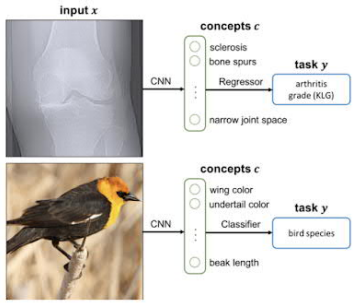
- VS the accuracy / interpretability tradeoff is a myth: https://www.youtube.com/watch?v=l0yrJz8uc5Q
  - In cases with "good data representation" (i.e. most models perform ~same)

252

252

Mixed Approaches

- E.g. concept bottleneck models
- Align predictions to pre-identified concepts
- Can increase interpretability
- Manual effort



Koh, P. W., Nguyen, T., Tang, Y. S., Mussmann, S., Pierson, E., Kim, B., & Liang, P. (2020). Concept Bottleneck Models. *ArXiv:2007.04612 [Cs, Stat]*. http://arxiv.org/abs/2007.04612

253

253

Different Types of Explainability

- Data explanation: understand the data
- Global direct explanation: the model itself is understandable
- Local direct explanation: an individual prediction is meaningful
- Global post hoc explanation: an understandable model explains the black box model
- Local post hoc explanation: an explanation is created for the individual prediction

Via: https://developer.ibm.com/blogs/explainable-ai-how-do-i-trust-model-predictions/

254

254

Emerging terminology…

- Interpretable
- Intelligible
- Interogatable
- Understandable

255

255

Interpretability

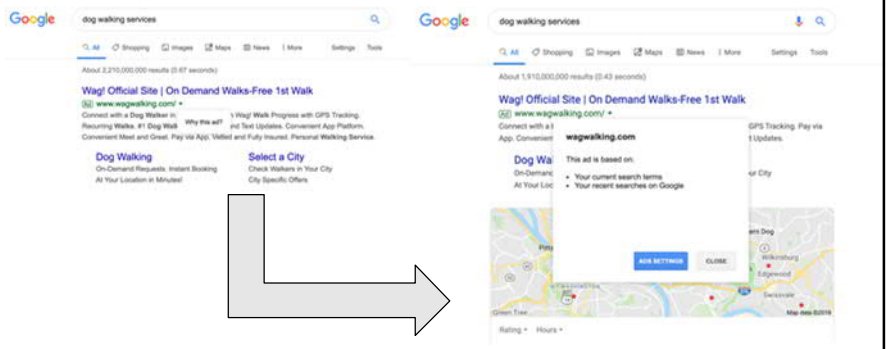
- Explainable (frequently): blackbox with posthoc explanations
- Interpretability (frequently): model that is not a black box
- Interpretability (sometimes): Focus on outputs

- x-able AI:
  - Relatable

256

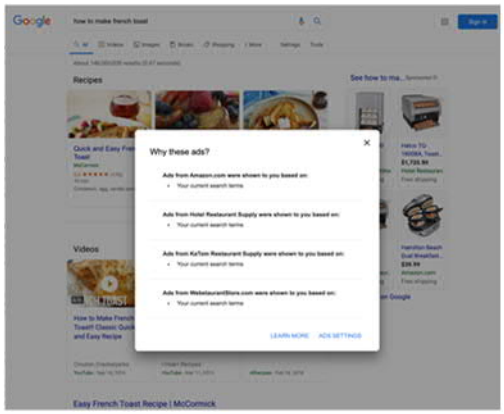
256

Good explanation/bad explanation?



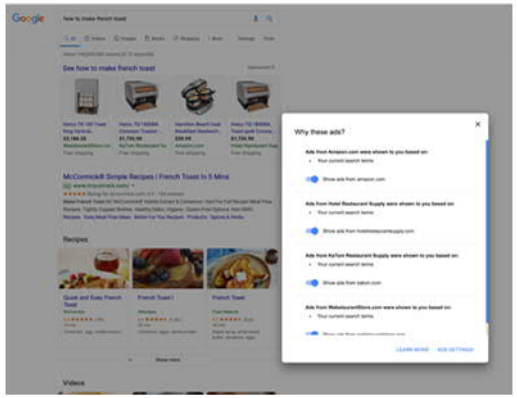
via https://haiicmu.github.io

Good explanation/bad explanation?



via https://haiicmu.github.io

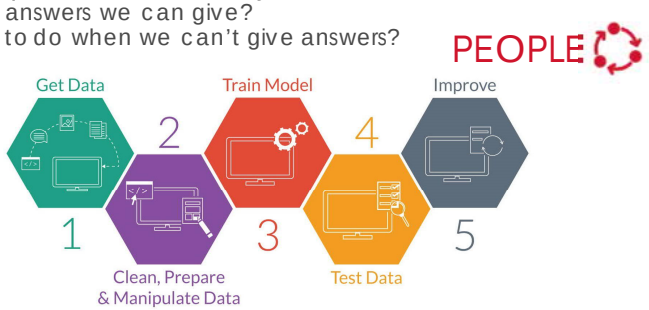
Adding Interaction (when logged in)...



via https://haiicmu.github.io

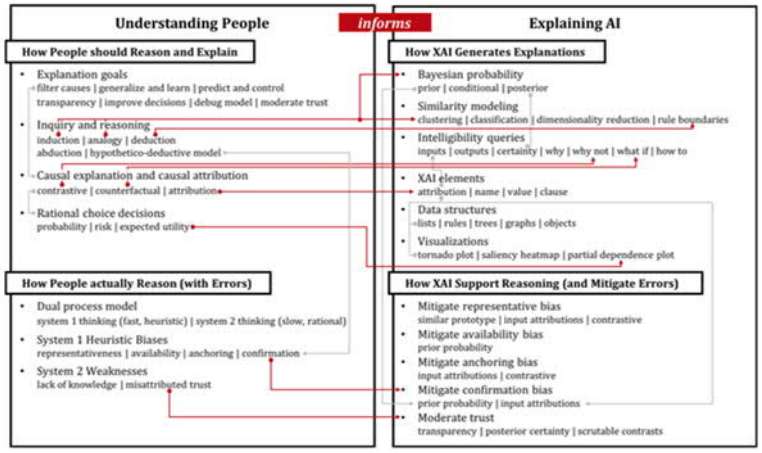
Human-Centred Approach

- What are the goals for different stakeholders?
- What questions could they have?
  - What answers we can give?
  - What to do when we can't give answers?



This Photo by Sunit Nandi is licensed under CC BY-SA-NC.

Cognitive Dimensions Approach



via: http://www.brianlim.net/wordpress/wp-content/uploads/2019/01/chi2019-reasoned-xai-framework.pdf

261

261

Questions and responses...

Possible "user questions":

1. Why did you do that?
2. Why not something else?
3. When do you succeed?
4. When do you fail?
5. When can I trust you?
6. How do I correct an error?

Hint: map questions to combinations of answers.

E.g. 1 -> [A+B]

Approaches to responses:

A. Simulation "run through decision process"
B. Decomposition "here's the information I used to make decision"
C. Algorithmic transparency: "Here's math that proves this algorithm will work, given large enough data"
D. Simplification: "My model is too complex to represent, but a simplified version is as follows..."
E. Examples: "Here are five other examples that are similar to this one that I classified the same way"

via https://haiicmu.github.io

262

262

Exploitability

- Abusive AI use / relationships
- Not talking about "hurting agents" (yet!?)...
- Embedding models can leak information
- Developer API access can be enough to be privacy critical
  - Esp. with models trained on private data
  - E.g. Thieves of Sesame Street: Model Extraction on BERT-based APIs
  - Extracted models quickly show high correlations with orig. mod.

263

263

The Case for Usable AI

- Modern AI in video games (as in many other application areas) usually far from state of the art
- Problem solving capacity vs. usability
- Critical "real-world use" considerations:
  - Plausibility / Believability
  - Computational Performance
  - Ease of Implementation

https://www.youtube.com/watch?v=aqXHsBrS6_U

Pfau, J., Smeddinck, J. D., & Malaka, R. (2020). The Case for Usable AI: What Industry Professionals Make of Academic AI in Video Games. In *Extended Abstracts of the 2020 Annual Symposium on Computer-Human Interaction in Play* (pp. 330–334). Association for Computing Machinery. https://doi.org/10.1145/3383668.3419905

264

264

Further Materials

- Clear (bit more detailed than this) intro to explainable AI: https://www.youtube.com/watch?v=AFC8yWzypss
- Provocation: Dangers of Explainability: https://www.youtube.com/watch?v=l0yfJ28uc5Q
- Overview of practical explainable AI (bit more technical perspective): https://www.youtube.com/watch?v=B-c8tIgchu0

This Photo by Admz is licensed under CC BY

265

Human-AI Interaction & Futures

Week 03 – Session 02: Ethics & AI

Jan Smeddinck

Unless otherwise noted, all materials: (CC BY-NC-SA 4.0)

266

266

Deep Fakes

- Impressive / fun ... but grave implications
- Can be & will be interactive

https://www.youtube.com/watch?v=VhFSlR7r7Yo

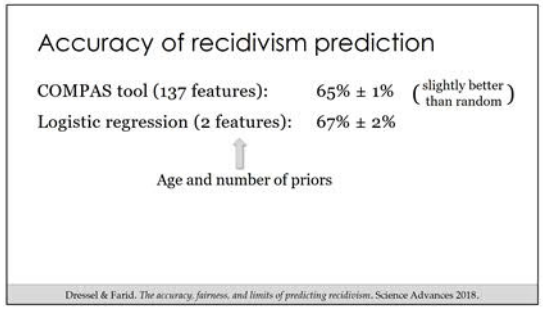
For more: "Ctrl Shift Face"
https://www.youtube.com/channel/UCKpH0CKltc73e4wh0_pgL3g

Do it yourself:
https://colab.research.google.com/github/AliaksandrSiarohin/first-order-model/blob/master/demo.ipynb

https://aliaksandrsiarohin.github.io/first-order-model-website/

267

Recidivism Prediction



Via: Narayanan, A. (n.d.). *How to recognize AI snake oil*. 21.

- Predicts re-arrest (!= recidivism)
- Reflects biases of policing
- Does not outperform traditional methods
- Obfuscates reasoning

See: https://advances.sciencemag.org/content/4/1/eaao5580

268

268

Complex Related Concepts

- Filter Bubbles and the Attention Economy
- Misinformation, Disinformation, Gossip and Bullshit … Consider: The Social Dilemma
- Not all AI-based, but can arguably intensify feedback-loops

269

269

AI-Solutionism

- Harms of AI for predicting social outcomes
- Hunger for personal data
- Massive transfer of power from domain experts & workers to unaccountable tech companies
- Lack of explainability
- Distracts from interventions
- Veneer of accuracy

Via: Narayanan, A. (n.d.). *How to recognize AI snake oil.* 21.

270

270

Both in Deployments & in Research

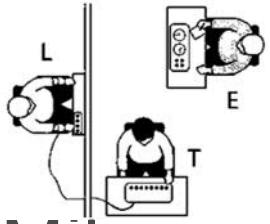
- 2020 Google scandal…
- Timnit Gebru (former) technical co-lead of Ethical AI
- "vs." Jeff Dean
- Forced or pressured resignation over AI Ethics paper
- See: https://www.theguardian.com/technology/2020/dec/04/timnit-gebru-google-ai-fired-diversity-ethics

271

This Photo by Kimberly White/Getty Images for TechCrunch is licensed under CC BY-SA

271

Psychology & Ethics



Milgram (1963)

272

272

Ethics History (Human Subject Research)

Mengele's twin „studies" (1940s)

Tuskegee syphilis exper. (1932 – 1972)

Image L: https://commons.wikimedia.org/wiki/File:Child_survivors_of_Auschwitz.jpeg (public domain)
Image R: https://en.wikipedia.org/wiki/File:Tuskegee_study.jpg (public domain)

People run these things ... often „following orders" ... troubling relation to locus of responsibility / justification practices w. AI

273

273

General Research Ethics

Following the European Code of Conduct for Research Integrity by the ESF (as of 2011)...

- honesty in communication;
- reliability in performing research;
- objectivity;
- impartiality and independence;
- openness and accessibility;
- duty of care;
- fairness in providing references and giving credit; and
- responsibility for the scientists and researchers of the future.

274

274

Ethical Key Concepts

- Act utilitarianism: A person's act is morally right if and only if it produces the best possible results in that specific situation
- Rule utilitarianism: A person's act is morally right if and only if it conforms to a *rule* that leads to the greatest good.
- Deontology: The morality of an action should be based on whether that action itself is right or wrong under a series of rules, rather than based on the consequences of the action.

275

275

Ethical Key Concepts

- Virtue ethics:
  - An act is moral if it is virtuous
  - A virtue is 'indeed a character trait—that is, a disposition which is well entrenched in its possessor, something that, as we say "goes all the way down"... ' It requires the *practically wise* agent.
    1. the practically wise agent's [has] capacity to recognise some features of a situation as more important than others, or indeed, in that situation, as the only relevant ones
    2. Practical wisdom can only come with experience (e.g. to know the likely consequences of certain actions for people)

→ *Current systems can't be blamed? But they will be? Accountability?*
Nearly any ethical frameworks disallows "I followed the rules" as defense.

276

276

Who is to blame?

"in March 2018, a self-driving Uber was navigating the Phoenix suburbs and failed to "see" a woman, hitting and killing her… In the case of Uber, the person minding the autonomous vehicle was ultimately blamed, even though Uber had explicitly disabled the vehicle's system for automatically applying brakes in dangerous situations."

via https://haiicmu.github.io from https://ainowinstitute.org/AI_Now_2018_Report.pdf

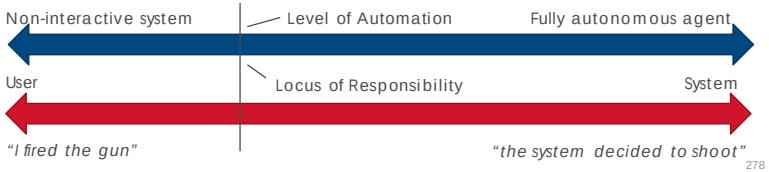
277

277

Responsibility & Justification Practices

- AI technology is problematic when it obscures who is making a decision → *accountability*
- Related work (around general interactive technologies) by Rosanna Bellini and Kieran Cutting at Open Lab

Bellini, R., Forrest, S., Westmarland, N., & Smeddinck, J. D. (2020). Mechanisms of Moral Responsibility: Rethinking Technologies for Domestic Violence Prevention Work. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–13. https://doi.org/10.1145/3313831.3376693



278

278

Fairness

Procedural Fairness: Decisions must be:

- Accurate
- Fair
- Consistent
- Correctable
- … (and other requirements)

Deals with *who gets how*. (cf deontological models)

- Bias / discrimination / etc.: fairness at individual, group, and societal level
- See also: https://t.co/hR9G9Bv6C7?amp=1

279

279

Procedural fairness leads to new AI requirements

Decisions must be:

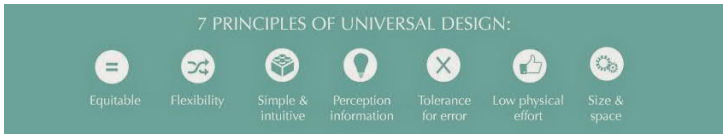
- Accurate
  - how do you deal with data unlike your training data? (credit scores for immigrants)
- Fair
  - the same outcome may be unfair (e.g. should you get extra late days on your assignment if your dog died?)
- Consistent
  - can a small change in your input lead to large differences in output?
- Correctable
  - What is our appeals process? Who/what decides the outcomes of the Final Appeal?

via https://haiicmu.github.io

280

280

Accessibility: Cf. Universal Design



This Photo by Unknown Author is licensed under CC BY-SA-NC

http://universaldesign.ie/What-is-Universal-Design/The-7-Principles/

281

281

Operability

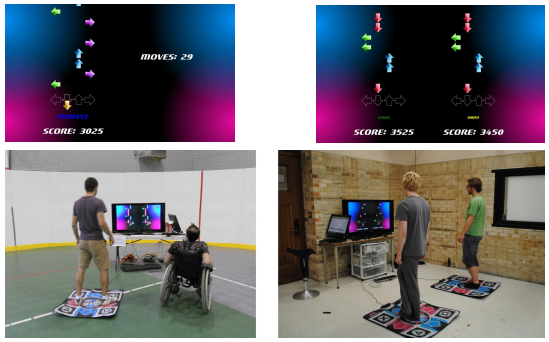
Everyone can use the system regardless of physical ability

- Minimise repetitive actions and need for sustained effort
- Use good affordances and constraints to make using controls
  easier (e.g. controls can only be used in the correct way)
- Support physical technologies (e.g. wheelchair access)
- Make sure that both seated and standing users can see information and controls.

282

282

Example: Co-Play and Balancing



Gerling, K. M., Miller, M., Mandryk, R. L., Birk, M. V., & Smeddinck, J. D. (2014). Effects of Balancing for Physical Abilities on Player Performance, Experience and Self-esteem in Exergames. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2201–2210. https://doi.org/10.1145/2556288.2556963

283

283

Further Considerations

- AI & Data Ethics Inseparable
- Complex relationships between ethics (focus on moral), policy & legal concerns, and commercial concerns
- Global-Local Problem
  - Diverse teams, weirdly spread out, global apps

284

284

Further Food for Thought

- AI Activism
  - Churchill, E. F. (2020). Reflecting on AI activism. *Interactions*, *28*(1), 26–27. https://doi.org/10.1145/3441563
  - Natarajan, S., & Parsheera, S. (2020). A conversation on AI activism. *Interactions*, *28*(1), 28–33. https://doi.org/10.1145/3436946
- Dual-Relationship
  - Ethics FOR AI!?! (when does consciousness arise? If it is ~ intelligence: do you want your Roomba to be conscious? Consequences if they are?)

285

285

Big Tech Frameworks

E.g.

- https://ai.google/principles
- https://www.microsoft.com/en-us/ai/responsible-ai?activetab=pivot1%3aprimaryr6
  - Fairness, Inclusiveness, Reliability & Safety, Transparency, Privacy & Security, Accountability

Review framework: https://arxiv.org/abs/2001.00973
Closing the AI Accountability Gap: Defining an End-to-End Framework for Internal Algorithmic Auditing

Model Cards for Model Reporting: Model Cards

286

286

Further Materials

- Recommended: Excellent quick and easy MOOC on AI Ethics: https://ethics-of-ai.mooc.fi/
  - AI systems raise ethical issues that include, but are not limited to, their impact on decision-making, equality, polarization, and well-being.
  - AI systems impact societal sectors such as employment and labor, social interaction, healthcare, education, weaponization, transport, and media.
  - AI systems cover topics such as freedom of expression, access to information, privacy, democracy, and discrimination.
  - AI systems can also change human experience, challenge human agency, raise concerns over the reliability of information sources, and question the ideal of fundamental dignity.
- General considerations on ethics in practice (in modern times): https://www.bitbybitbook.com/en/1st-ed/ethics/ (Links to an external site.)

287

287

Human-AI Interaction & Futures

Week 03 – Session 03:
Humans-in-the-Loop

Jan Smeddinck

Unless otherwise noted,
all materials: (CC BY-NC-SA 4.0)

288

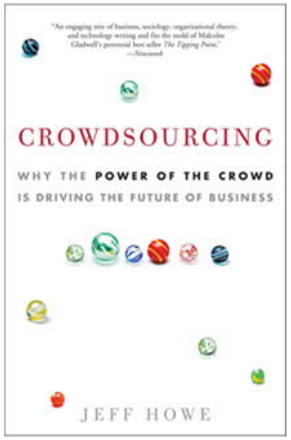
288

Human → ←AI

- Humans can be (more or less) active contributors, not "just consumers"
- Humans can be involved (beyond being researchers and developers) in any step of the AI lifecycle process...
  - E.g. collect data, label examples, outcome validation, etc.
- Key concepts: *crowdsourcing & human computation*

289

289



"Crowd" + "Outsourcing"

- One specific area of application that leverages wisdom (or, at the very least, earnest contributions) of crowds
- Originally used to describe how businesses use the Internet to "outsource work to the crowd"
- Now also used to describe how many different types of online projects (including academic and non-profit) use crowds for social good.

290

Human Computation (HComp)

- Early: ESP Game
- Later: Captcha, Duolingo

Ahn, L. von. (2007). Human computation. *Proceedings of the 4th International Conference on Knowledge Capture*, 5–6. https://doi.org/10.1145/1298406.1298408

Led to e.g. MTurk ...

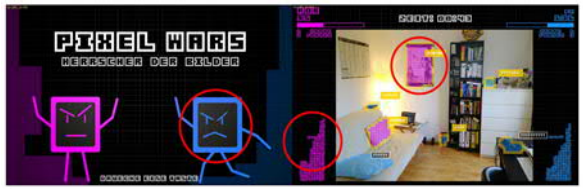
- How to present (break down?) problems? Task difficulty?
- How to collect data?
- How to fuse data?
- Data validation?
- Motivation?
- Ethics? (crowdwork / gig-work in general)

291

Motivation

EXTRINSIC
- Money
- Reputation
- Awards
- Networking

INTRINSIC
- Fun (e.g. games)
- Altruism
- Social comparison / compet.
- Learning



Motivierende Effekte verschiedener Level der Gamification im Bereich Human-based Computation für Image Labeling und Segmentierung (Robin Wienschendorf, 2017)

292

Human Computation Games: Robots & Pancakes

https://youtu.be/REEgBzvcjMQ

Walther-Franks, B., Smeddinck, J., Szmidt, P., Haidu, A., Beetz, M., & Malaka, R. (2015). Robots, Pancakes, and Computer Games: Designing Serious Games for Robot Imitation Learning. *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, 3623–3632. https://doi.org/10.1145/2702123.2702552

293

293

Quality Control

Imagine that you're setting up a crowdsourcing task for image classification: "Dog or Muffin?"

What "quality control" strategies would you use to get an accurate answer for each image?

via https://haiicmu.github.io

294

2. Quality control

Common mechanisms:

- Tutorials
- Redundancy
  - Voting
  - Discussion

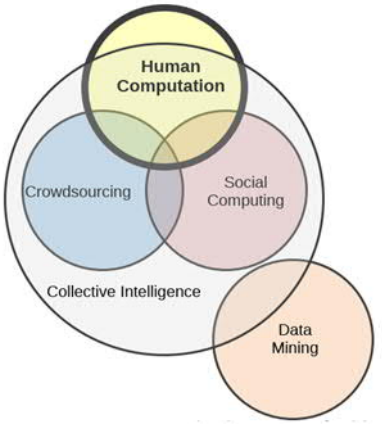
via https://haiicmu.github.io

295

Taxonomy

- Human computation: means of solving computational problems
- Rarely found in crowdsourcing / social computing applications

Quinn, A. J., & Bederson, B. B. (2011). Human computation: A survey and taxonomy of a growing field. *Proceedings of the 2011 Annual Conference on Human Factors in Computing Systems*, 1403–1412.



296

296

Problem Spaces for HComp

- Intuitive Decisions
  - Now being partially replaced (e.g. in perception)
  - Still holds as intuitive judgment / in social contexts
- Aesthetic Judgment
- Contextual Reasoning
- Embodiment Issues

Krause, M., & Smeddinck, J. (2011). Human Computation Games: A Survey. *19th European Signal Processing Conference, 2011*, 754–758.

297

297

Some roles for humans

- Data feeders: when machines can't read this data, but humans can.
- Backup: Machines do the easy tasks, humans do the hard stuff
- Triagers: Humans see which particular task needs to be done by people, which one can be automated (or deciding between different automation methods)
- Appeals judge: Humans override the algorithm when it's wrong
- Worker: Human does arbitrary task under machine supervision

via https://haiicmu.github.io

298

298

Some roles for machines

- Advisor: suggest (but not mandate) actions for one person
- Orchestrator: suggest (but not mandate) actions for many people
- Questioner: ask humans to consider/re-think some aspects of the task
- Manager: Tell humans what to do in task
- Lifeguard: Prevent humans from making high-cost mistakes
- Worker: Do specific aspects of the task that humans direct

via https://haiicmu.github.io

299

299

Humans in the Loop with AI and ML

- From data generation / curation to "live function calls"
- In ML e.g.: *active learning*

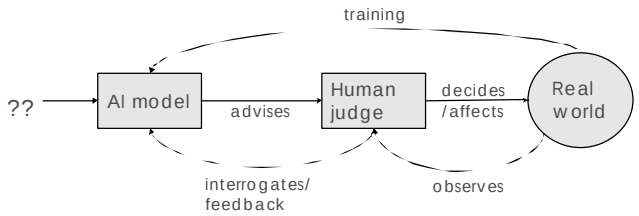
- Example for automated exercise execution renderings based on text:
  Sarma, H., Baran Samaddar, A., Porzel, R., D. Smeddinck, J., & Malaka, R. (2017). Updating Bayesian networks using crowds. *Neural Network World*, *27*, 529–540. https://doi.org/10.14311/NNW.2017.27.028

- Example from self-driving cars: https://drive.google.com/file/d/1htnA4_bUdfXdpDo033Tz0NLjMGG_wfuK/view

300

300

Advice-giving AI: How might bias creep in?



via https://haiicmu.github.io

301

301

Further Materials

- HComp along examples (esp tied to ML): https://www.youtube.com/watch?v=5ymW5War-P8
- HComp via the story of Duolingo (consider: still relevant for translations?), by Mr HComp in persion (Luis van Ahn): https://www.youtube.com/watch?v=cQI6jUjFjp4
- https://hai.stanford.edu/blog/humans-loop-design-interactive-ai-systems

302

302

Human-AI Interaction & Futures

Week 03 – Session 04: Recommender Systems

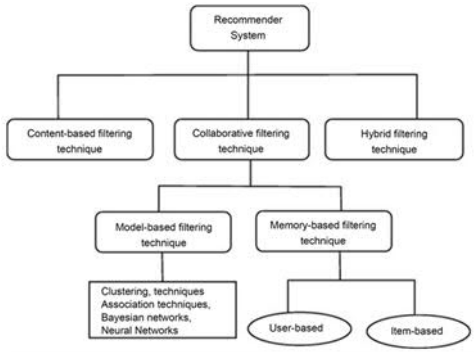
Jan Smeddinck

Unless otherwise noted, all materials: (CC BY-NC-SA 4.0)

303

303

Applications: Recommender Systems



Mylavarapu, B. K. (2018). Collaborative Filtering and Artificial Neural Network Based Recommendation System for Advanced Applications. *Journal of Computer and Communications*, *6*(12), 1–14. https://doi.org/10.4236/jcc.2018.612001

304

304

Applications: Recommender Systems

- Collaborative filtering

	i_1	i_2	i_3	i_4	i_5	...	i_{32}	
Jack	5	1	3	4	3	...	1	similar users based on profile
u_2	0	4	0	0	0	...	5	
u_3	0	2	0	0	0	...	4	
Tom	4	1	4	5	4	...	2	
...	...	...	...	...	...	...	...	
Myself	5	1	4	?	3	...	1	target user

predicted rating: 4.5

This Photo by WikiBooks is licensed under CC BY-SA

305

305

Recommender Systems

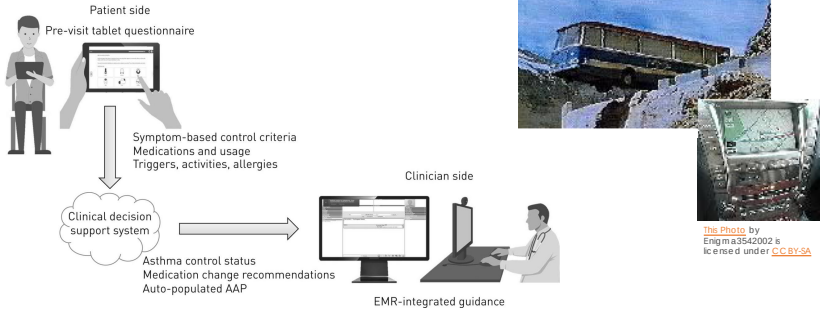
- What if you see an entirely new user?
  - Avoid predicting all 0 due to optimisation terms .. Mean normalisation…
- Also content-based recommendations
- Can be formulated as supervised learning problem
  - Optimisation via gradient descent

306

306

Complex Recommender Systems

- Decision support



This Photo by Enigma3542002 is licensed under CC BY-SA

Gupta, S., Price, C., Agarwal, G., Chan, D., Goel, S., Boulet, L.-P., Kaplan, A. G., Lebovic, G., Mamdani, M., & Straus, S. E. (2019). The Electronic Asthma Management System (eAMS) improves primary care asthma management. *European Respiratory Journal*, *53*(4).

307

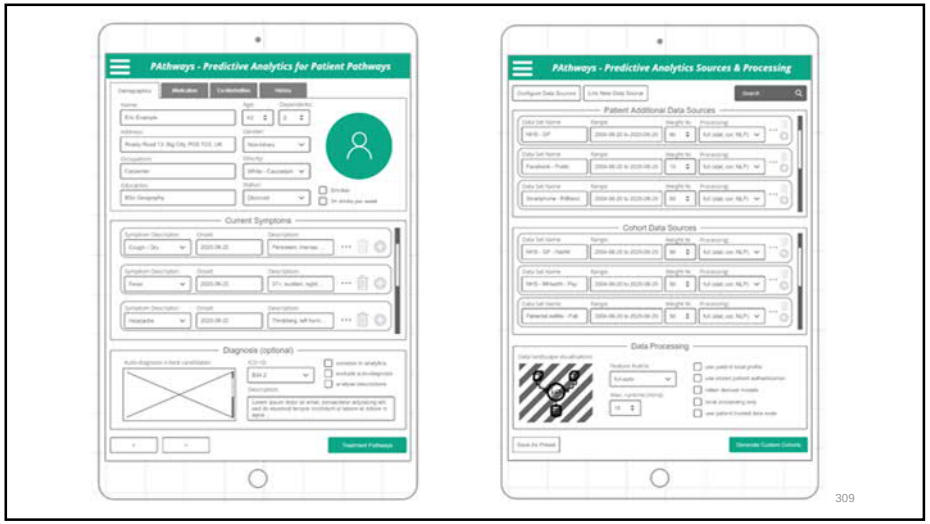
307

Decision Support Systems

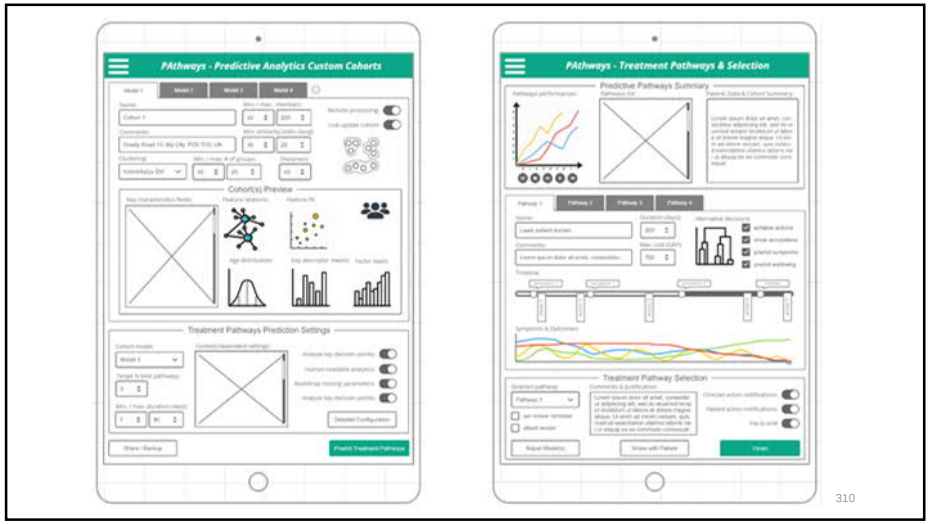
- E.g. in digital health:
  Zhu, N., Cao, J., Shen, K., Chen, X., & Zhu, S. (2020). A Decision Support System with Intelligent Recommendation for Multi-disciplinary Medical Treatment. *ACM Transactions on Multimedia Computing, Communications, and Applications*, *16*(1s), 33:1-33:23. https://doi.org/10.1145/3352573
- Diagnostics, ICD-codes, most viable mediation(s), etc.

308

308



309



310

Additional Learning Materials

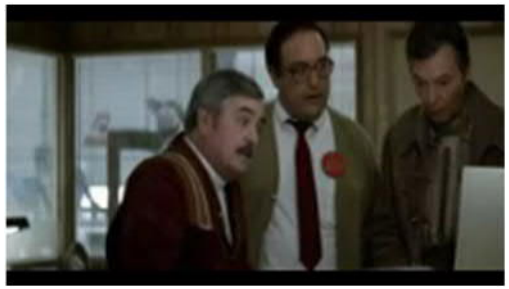
- Slower version intro: https://www.youtube.com/watch?v=Eeg1DEeWUjA
- More thorough and technical intro by Andrew Ng himself: https://www.youtube.com/watch?v=gilXNoiqO_U&list=PLLssT5z_DsK-h9vYZkQkYNWcItqhlRJLN&index=96 (16.1-16.6)

311

311

312

Hello Computer



https://www.youtube.com/watch?v=hShY6xZWVGE

313

313

Conversational Desktop (1987)

https://www.youtube.com/watch?v=HGrkpvy4k3Q

314

314

The Coordinator (1987)

CONVERSE

OPEN CONVERSATION FOR ACTION	REVIEW / HANDLE
Request	Read new mail
Offer	Missing my response
	Missing other's response
OPEN CONVERSATION FOR POSSIBILITIES	
Declare an opening	My promises/offers
	My requests
ANSWER	Commitments due: 24-May-88
NOTES	Conversation records

conversation moves towards mutual agreement

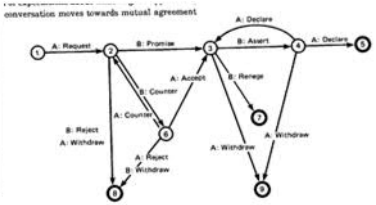


Figure 1. The Basic Conversation for Action.

SPEAKING IN A CONVERSATION FOR ACTION

Acknowledge	Promise
Free-Form	Counter-offer
Commit-to-commit	Decline
Interim-report	Report-completion

email and other communication

http://hci.stanford.edu/~winograd/papers/language-action.html

Winograd, T. (1986). A language/action perspective on the design of cooperative work. *Proceedings of the 1986 ACM Conference on Computer-Supported Cooperative Work*, 203–220. https://doi.org/10.1145/637069.637096

315

315

Interaction: Natural language

This Photo by ZoxZoxZ is licensed under CC BY-SA

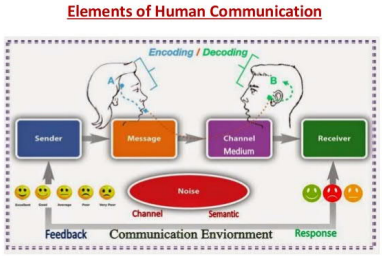
- NATURAL
- speech or text
- vague
- ambiguous
- try a subset e.g. single commands
- Recent boom (chat and voice based)

Example: Search engines, help systems, phone helplines

316

316

Cognitive Dimensions: Communication



https://www.slideshare.net/zahraafat/human-communication-61657557

https://www.flickr.com/photos/63168233@N06/8215739815

317

317

Compression & Aliasing

- Discretisation / Quantisation
- Sampling
- Aliasing (rasterization & signal undersampling / oversampling)
- Nyquist Theorem
  (sample at 2x frequency for which aliasing should be avoided)
- Lossless vs. Lossy (perceptive)

- More detailed (but still quick) explanation:
  https://www.youtube.com/watch?v=yWqrx08UeUs

318

318

Signifiers / NLI

- Forms
- Question-answer systems
- Query-language
- Text-based or spoken
- Natural language interfaces
  - Build on natural language processing (NLP)
    - Speech recognition
    - Speech synthesis

Charles Sanders Peirce, semiotician, look him up about the importance of context for meaning!

319

319

Metaphors

- Figure of speech
- Directly refers to one thing by mentioning another
- May provide (or obscure) clarity
- May identify hidden similarities between two ideas
- Very important in interaction design (e.g. "the desktop metaphor")
- Can be used in human-AI interaction (e.g. visual agents design), but also may be accidentally implied…

- Take a look at "Metaphors we live by" (George Lakoff)

320

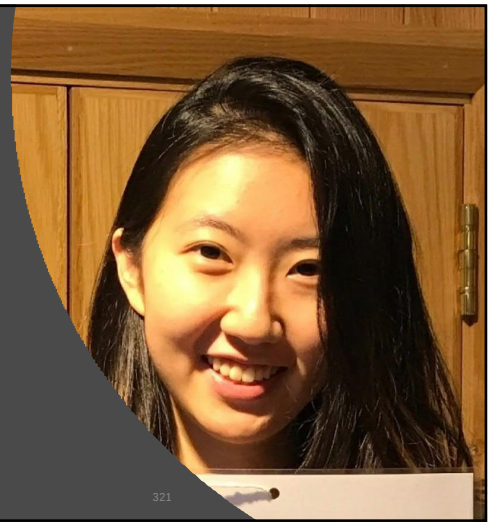
320

Dialogical Service Provisioning

From ongoing work with Viana (Nijia) Zhang (supervisee at Open Lab):

- Maternal mental health and well-being
- Un-platforming approach / chat-based

→ *Focus: NLI aspects*



N.Zhang10@newcastle.ac.uk

321

Conv. Interfaces ←→ Conv. Agents

- Language != conversation
- Dialogue (social rules implied)
- Arguably building block of cognition
  - See Wittgenstein: Language Games
- → *(some degree of) antropomorphologisation*

See e.g.: Wang, Q., Saha, K., Gregori, E., Joyner, D. A., & Goel, A. K. (2021). *Towards Mutual Theory of Mind in Human-AI Interaction: How Language Reflects What Students Perceive About a Virtual Teaching Assistant*. 15. http://qiaosiwang.me/Publications/MToM_Preprint.pdf

322

Modern AI for NLP/NLI

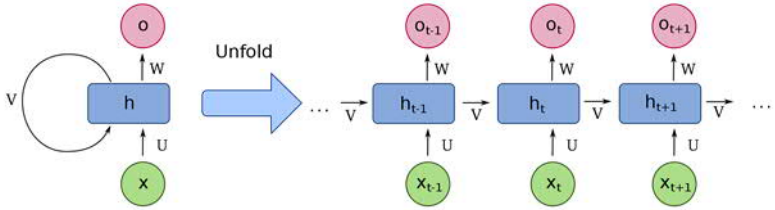
Language unfolds over time … need to allow methods to capture that…

- Recurrent Neural Network (RNNs)
- Long Short Term Memory RNN (LSTM)
- Transformers

See also:
https://towardsdatascience.com/recurrent-neural-networks-deep-learning-for-nlp-37baa188aef5
And: https://medium.com/analytics-vidhya/natural-language-processing-from-basics-to-using-rnn-and-lstm-ef6779e4ae66

323

Recurrent Neural Networks



This Photo by Unknown Author is licensed under CC BY-SA

324

LSTM: Long Short Term Memory RNN

LSTM adds 3 gates to manage memory

The Gate of Forgetting

What past hidden state is worth keeping (still relevant) or no: (0 forget - 1 keep) * past

The Gate of Input

What past hidden state will be useful to figure out this input? Input + (remembered past)

The Gate of Output

What out of past state + this current input will be useful later?

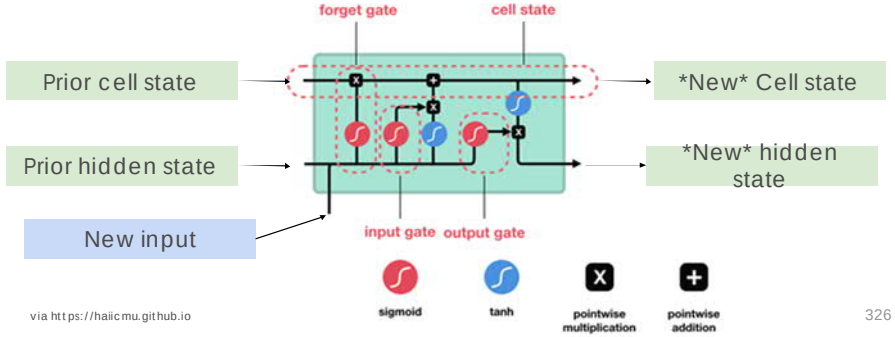
via https://haiicmu.github.io

325

325

LSTM model (a neuron is now called a cell)

LSTM also adds a long term memory (cell state) to complement a more short term memory (hidden state).



via https://haiicmu.github.io

326

326

How well does this perform?

327

327

Fun with RNNs / LSTMs (2016)

http://www.youtube.com/watch?v=LY7x2Ihqjmc

"Sunspring" is a short film who's script was written by an LSTM. What do you notice about the story structure?

via https://haiicmu.github.io

328

328

Memory and attention

What are some things this voice agent (HAL) is able to do that our voice agents are not able to do today?

2001: A Space Odyssey http://www.youtube.com/watch?v=r13I-TuDcWI

via https://haiicmu.github.io

329

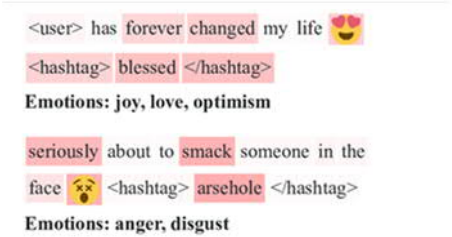
329

In recalling sequences, what information do you pay attention to?

330

330

Neural Nets with Attention



Like in this heatmap visualization, "Attention" is a technique that can be used to see what words in a sequence the model pays the most attention to make its prediction.

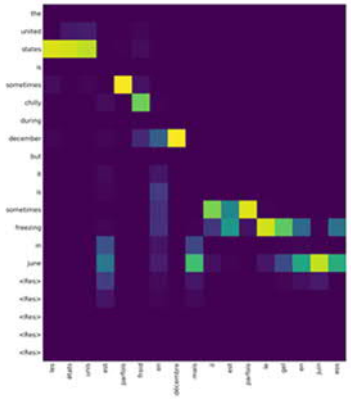
Shown is sentiment analysis for two Tweets.

via https://haiicmu.github.io

331

331

Neural Nets with Attention



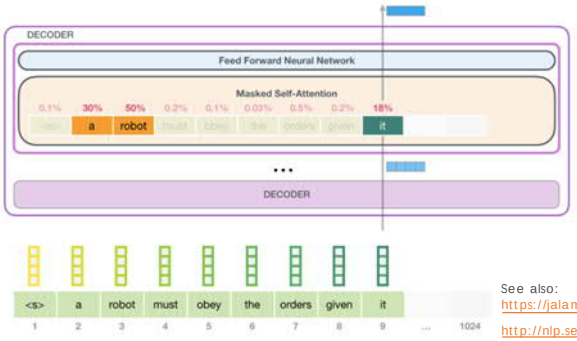
Like in this heatmap visualization, "Attention" is a technique that can be used to see what words in a sequence the model pays the most attention to make its prediction.

Shown is attention for translating a phrase from english to french

332

332

Transformers: NNs w. Attention (2019)



Intuition: aggregating hidden state across long sequences can't account for long spaced out dependencies.

Instead pass hidden state after each word, and use attention to figure out which 'chunk' of history is most important at any given point.

See also:
https://jalammar.github.io/illustrated-transformer/

http://nlp.seas.harvard.edu/2018/04/03/attention.html

Via https://jalammar.github.io/illustrated-gpt2/

GPT-3: Attention Focus & Billions of Parameters

What It's Like To be a Computer: An Interview with GPT-3:
https://www.youtube.com/watch?v=PqbB07n_uQ4

https://ai.googleblog.com/2017/08/transformer-novel-neural-network.html

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., ... Amodei, D. (2020). Language Models are Few-Shot Learners. *ArXiv:2005.14165 [Cs]*. http://arxiv.org/abs/2005.14165


A "whatpu" is a small, furry animal native to Tanzania. An example of a sentence that uses the word whatpu is:
We were traveling in Africa and we saw these very cute whatpus.

To do a "farduddle" means to jump up and down really fast. An example of a sentence that uses the word farduddle is:
One day when I was playing tag with my little sister, she got really excited and she started doing these crazy farduddles.

A "yalubalu" is a type of vegetable that looks like a big pumpkin. An example of a sentence that uses the word yalubalu is:
I was on a trip to Africa and I tried this yalubalu vegetable that was grown in a garden there. It was delicious.

A "Burringo" is a car with very fast acceleration. An example of a sentence that uses the word Burringo is:
In our garage we have a Burringo that my father drives to work every day.

A "Gigamuru" is a type of Japanese musical instrument. An example of a sentence that uses the word Gigamuru is:
I have a Gigamuru that my uncle gave me as a gift. I love to play it at home.

To "screeg" something is to swing a sword at it. An example of a sentence that uses the word screeg is:
We screeghed at each other for several minutes and then we went outside and ate ice cream.

Figure 3.16: Representative GPT-3 completions for the few-shot task of using a new word in a sentence. Boldface is GPT-3's completions, plain text is human prompts. In the first example both the prompt and the completion are provided by a human; this then serves as conditioning for subsequent examples where GPT-3 receives successive additional prompts and provides the completions. Nothing task-specific is provided to GPT-3 other than the conditioning shown here.

Fun with Transformers (2019)

This Photo by Unknown Author is licensed under CC BY-SA

A text model based on a model released by OpenAI in early 2019. What do you think of the quality?

https://talktotransformer.com/

or

https://transformer.huggingface.co/doc/distil-gpt2

Crossing Domains

- Transformers
- originally developed for language problems
- Growing body of work on other domains:
  - Images
  - Videos
  - Speech
  - protein folding
  - automated coding (sort of language)
  - ...

Human-AI Interaction & Futures

Week 03 – Session 06:
Agents, Bots & Robots

Jan Smeddinck

Unless otherwise noted,
all materials: (CC BY-NC-SA 4.0)

Learning Goals

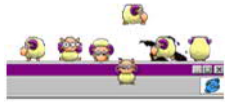
Broad understanding of interaction concerns around:

- Bots & (interface) agents
- Embodied conversational agents & NPCs
- Robots & human-robot interaction

See: https://telegram.me/botfather

Bots

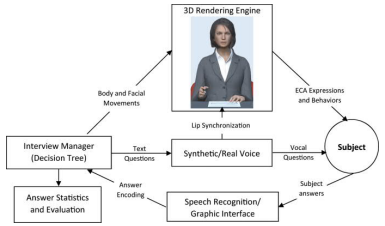
- Most bots are simply apps that run scripts on the internet (aka web robots aka Internet bots)
  - E.g. search engine spiders (interaction: robots.txt)
  - Many are "bad" … e.g. "spambots"
- Often "chatbots" (i.e. conversational agents)
  - Early forms in IRC



This Photo by текущий is licensed under CC BY

Embodied Conversational Agents (ECAs): OLGA (1998)

- Chatbots with (virtual) "bodies"
- May / may not have voice
- May / may not use text



This Photo by Unknown Author is licensed under CC BY

https://www.youtube.com/watch?v=Rt9oKtosMig

ECAs in MR: Welbo (2000)

DOI:: https://doi.org/10.1145/633292.633299
WEB:: http://www.cs.umd.edu/hcil/chivideosl...

Anabuki, M., Kakuta, H., Yamamoto, H., & Tamura, H. (2000). Welbo: An embodied conversational agent living in mixed reality space. *CHI '00 Extended Abstracts on Human Factors in Computing Systems*, 10–11. ACM. https://doi.org/10.1145/633292.633299

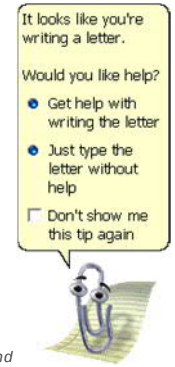
https://www.youtube.com/watch?v=Esh9f1SSD1s

341

341

What killed early ECAs?

- E.g. MS Office Assistant used Bayesian Nets tech to offer support ... yay!?
- Most generally anthromorphopologisation ...
  - I.e. "it looks and acts at if it might be smart, so it better be smart"
  - Projected abilities often misconceptions
  - Complex rules of social behaviour / etiquette
  - Complex "rules" of dialogue
  - Some people actually liked Clippit (et al.)



Swartz, L. (2003). *Why People Hate the Paperclip: Labels, Appearance, Behavior, and Social Responses to User Interface Agents.* https://doi.org/10.13140/RG.2.1.2508.1047

342

342

ECAs

- Mid 90s to early 00s wave ... but persistent issues
- Resurgence since 2010+
- Never went away, esp. through "cousin" development of non-player characters (NPCs) in video games

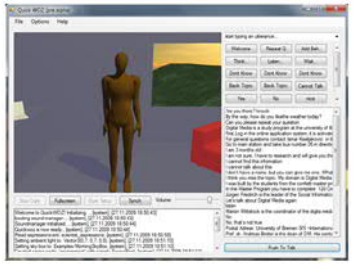


Figure 1. Screenshot of a QuickWoZ sample scene with quick-play buttons on the top right.

Dale, R. (2016). The return of the chatbots. *Natural Language Engineering*, *22*(5), 811–817. https://doi.org/10.1017/S1351324916000243

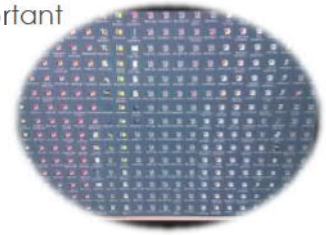
Smeddinck, J., Wajda, K., Naveed, A., Touma, L., Chen, Y., Hasan, M. A., Latif, M. W., & Porzel, R. (2010). QuickWoZ: a Multi-purpose Wizard-of-Oz Framework for Experiments with Embodied Conversational Agents. *Proceeding of the 14th International Conference on Intelligent User Interfaces*, 427–428. http://doi.acm.org/10.1145/1719970.1720055

343

343

Future – DM vs. Interface Agents

- Direct Manipulation will stay important
- Become more intuitive
- Faster with better feedback
- Limits with huge data sets
  - Agents can help handling that
  - or: SoundTorch
    - http://vimeo.com/groups/digmed/videos/2446867



- Be polite to users and give them controls that let them focus on their jobs

Shneiderman, B., & Maes, P. (1997). Direct manipulation vs. Interface agents. *Interactions*, *4*(6), 42–61. https://doi.org/10.1145/267505.267514

http://farm1.static.flickr.com/8/7157243_a5cf2853bc.jpg?v=0

344

344

This Photo by Unknown Author is licensed under CC BY

This Photo by Matias Lenin is licensed under CC BY-SA

345

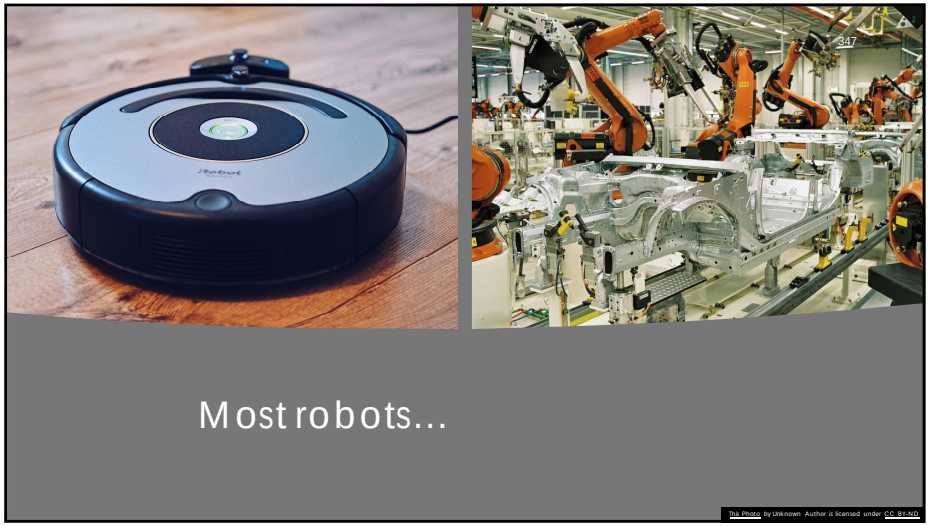
345

346

Not an abstract future…

- … also:
- Robert Williams (May 2, 1953 – Jan. 25, 1979)
  - Killed by industrial robot arm at Ford Motor factory https://en.wikipedia.org/wiki/Robert_Williams_(robot_fatality)

346



347

Most robots…

This Photo by Unknown Author is licensed under CC BY-ND

347

ECAs & Robots in the Consumer Market

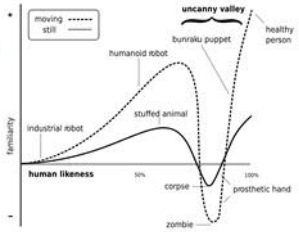
Or “inspired products” … depending on who you ask.

348

349

Uncanny Valley

- Since we are talking about Boston Dynamics:
- https://www.youtube.com/watch?v=YdnJI9T-yXI
- This can apply to motion dynamics but first discussed about static visual appearance...
- https://www.youtube.com/watch?v=GVdx22WtXI4
- More scary (did anybody plan on sleeping this month?):
- https://www.youtube.com/watch?v=IM82RzN0urI



This Photo by Unknown Author is licensed under CC BY-SA

Seyama, J., & Nagayama, R. S. (2007). The uncanny valley: Effect of realism on the impression of artificial human faces. *Presence: Teleoperators and Virtual Environments*, *16*(4), 337–351.

350

350

Interaction → Coordination → Collaboration

Baxter https://www.youtube.com/watch?v=uLTBejGnxdE

351

351

Evolutionary Linguistics & Embodied Language Games

Luc Steels language development in robots: https://www.youtube.com/watch?v=Qh2yT-AL1V8

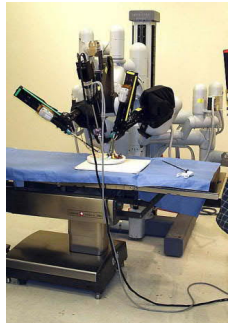
352

352

Robots in (e.g.) Health Applications

Wood, L. J., Zaraki, A., Robins, B., & Dautenhahn, K. (2019). Developing Kaspar: A Humanoid Robot for Children with Autism. *International Journal of Social Robotics*. https://doi.org/10.1007/s12369-019-00563-6

Social robots in elderly care:
https://www.youtube.com/watch?v=ppPLDEi82Ig



This Photo by Unknown Author is licensed under CC BY-SA

Fard, M. J., Ameri, S., Ellis, R. D., Chinnam, R. B., Pandya, A. K., & Klein, M. D. (2018). Automated robot-assisted surgical skill evaluation: Predictive analytics approach. *The International Journal of Medical Robotics and Computer Assisted Surgery, 14*(1), e1850. https://doi.org/10.1002/rcs.1850

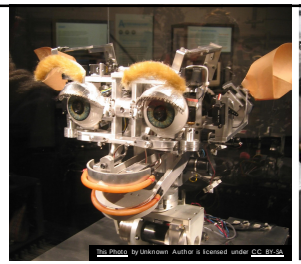
353

353

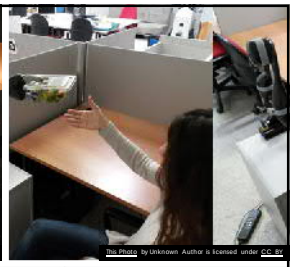
Human-Robot Interaction

From Asimov's robot laws to own sub-field...

- Perceiving humans
- Motion / action planning
- Cognitive models / theory of mind
- Human-robot coordination / collaboration



This Photo by Unknown Author is licensed under CC BY-SA



This Photo by Unknown Author is licensed under CC BY

This Photo by Unknown Author is licensed under CC BY-NC

354

354

A revolution in the making (2020) ...

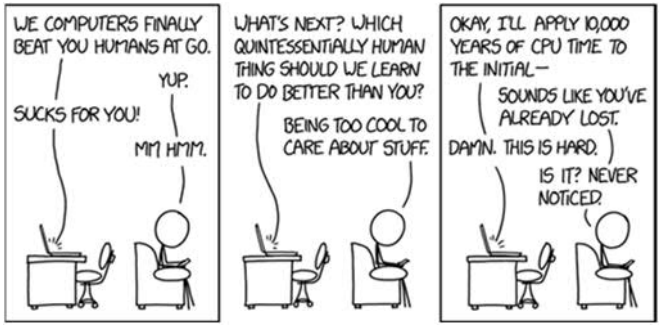
https://www.youtube.com/watch?v=fn3KWM1kuAw

See also: Spot with an arm:
https://www.youtube.com/watch?v=WvTdNwyADZc%2F&t=1120

355

355

Not to worry... (daily XKCD)



Further materials (free ebook): https://www.human-robot-interaction.org/

356

356

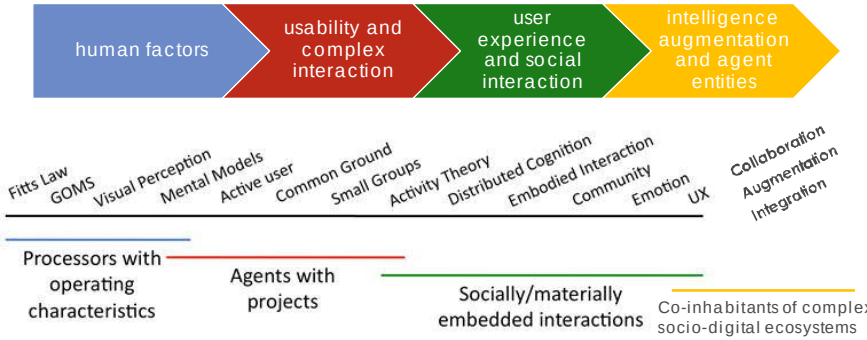
Human-AI Interaction & Futures

Week 03 – Session 07:
Human-AI Integration

Jan Smeddinck

Unless otherwise noted, all materials: (CC BY-NC-SA 4.0)

History of HCI + AI ???



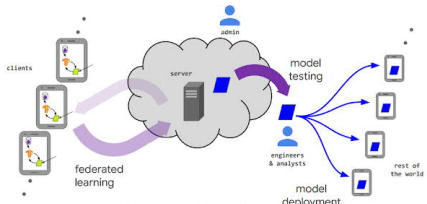
28/12/2021

via: https://www.interaction-design.org/literature/book/the-encyclopedia-of-human-computer-interaction-2nd-ed/human-computer-interaction-brief-intro

Human-AI Relationships

- Missing medium- to long-term perspectives...
- ... what does it do to/with/for humans?
- ... how to design interactions?
- ... what can be gained from a system / learning perspective?
  - Recent trend in federated learning ...

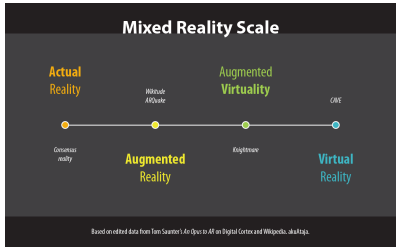
The lifecycle of an FL-trained model and the various actors in a federated learning system:



See: federated learning

Mixed Reality

- ... and now it is getting complicated ... https://www.youtube.com/watch?v=YJg02ivYzSs



This Photo by Unknown Author is licensed under CC BY-SA

Direct Manipulation → Mind Control

- Vibrovests: remote control your dog: https://www.youtube.com/watch?v=ofNjevpHdX0
- EMS – Remote control friends
- https://www.youtube.com/watch?v=JSfnm_HoUv4

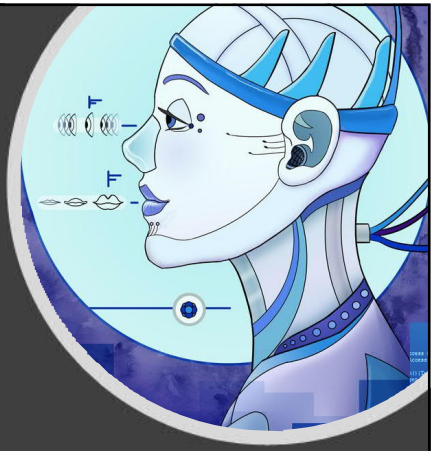
- BCIs: for telepathy and (mediated) telekinesis https://www.youtube.com/watch?v=kR1wvi2EFxA
- TMS: remotely…
- Mind reading: https://www.youtube.com/watch?v=lUg-t609byg

361

361

Transhumanism

- CCS-HCI
- Homo sapiens BRANCH with homo optimus?
- How will we communicate / translate?



362

362

Curiosities → Pets → Workers → Collaborators

- Nearly there …
- … and then what?

- Which robot will be the "iPhone of robotics"?
  - Application area?
  - Capabilities?
  - Interaction modalities?

This Photo by Unknown Author is licensed under CC BY-SA-NC

363

363

Human-AI Interaction & Futures

Week 03 – Session 07:
Creative AI &
Human-AI Co-Creation

Jan Smeddinck

Unless otherwise noted,
all materials: (CC BY-NC-SA 4.0)

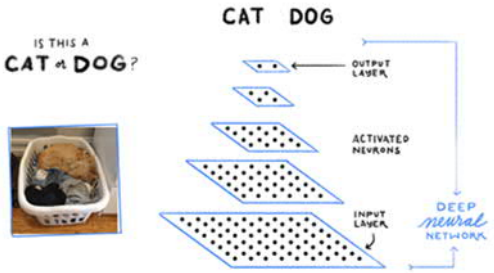
364

364

“Simple” object recognition

Applications include:

- Sorting and indexing your photos
- Gathering and labeling data
- What else?



via https://haiicmu.github.io

Face recognition (Facebook 2019)

Facebook $35B class action lawsuit. Why?

The state of Illinois has a law that individuals have to consent to having their biometrics (including face) collected and scanned by AI.

$35B = $1-5000 per 7 million users in Illinois

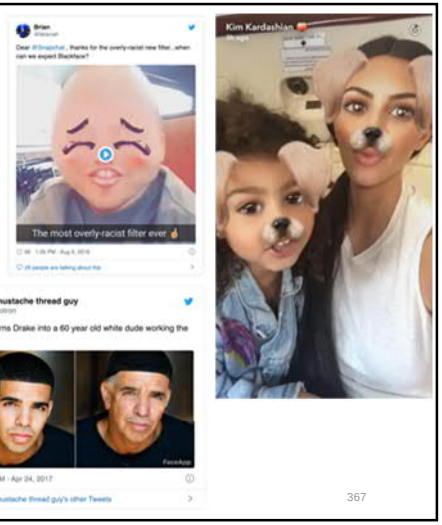
2021 update: settlement $300 each to 1.6M users ..

via https://haiicmu.github.io

Image Fun!

Image filter games and apps make great social media and meme material.

- Privacy implications?
- Racial implications?



via https://haiicmu.github.io

Human Deep Fakes (2018)

[Warning:
video contains harsh political satire]

This video helped set off a bit of a public scare around deep fakes.

https://www.youtube.com/watch?app=desktop&v=cQ54GDm1eL0

via https://haiicmu.github.io

Style transfer

Neural Net is trained to learn a particular style.

We use transfer learning to apply that learned style to a new image to stylize it.

More examples:
https://deepart.io/latest/

via https://haiicmu.github.io

369

369

CNN: Convolutional Neural Net

CNNs are a type of multilayer perceptron most commonly used for image tasks.

We add a concept of "memory" to process serial data.
What kind of abilities might a neural net need for vision?

Croatian sheepdog | Costa Rica stray dog estate | Herding pup in training

via https://haiicmu.github.io

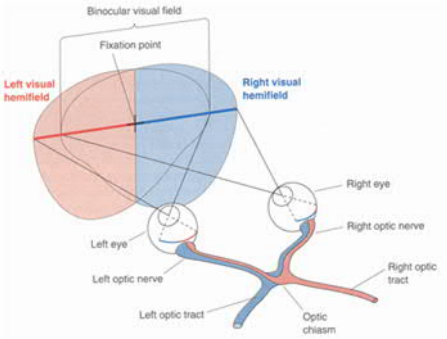
370

370

Human visual perception

CNNs are inspired by visual cortex:

1. We take in an "image" as what we see in our full visual receptive field.
2. Break down an image into subparts.
3. Each visual neuron is responsible for a small part (or aspect) of the receptive field.
4. Neurons *partially overlap* their receptive field. What is the benefit of doing so?
5. Higher layers of neurons combine the info from lower layers to perceive abstract visual information ("cat")



via https://haiicmu.github.io

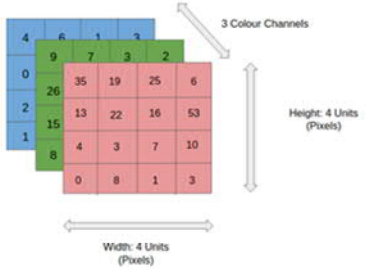
via https://haiicmu.github.io 371

371

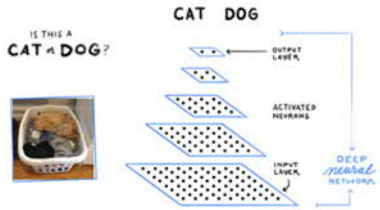
372

CNN building blocks: Input and Output Layers

Input Layer will be e.g. a colored RGB image



Output layer will be a classification of the image



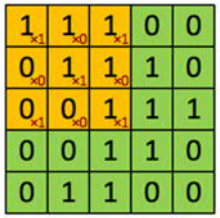
via https://haiicmu.github.io

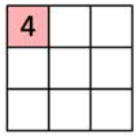
372

CNN building blocks: Convolution

In yellow, there is a 3x3 convolution with a stride (movement) of 1.

- The yellow is where a neuron is "looking" at the image
- The pink convoluted feature is a summary of what each neuron "sees" in its yellow receptive field

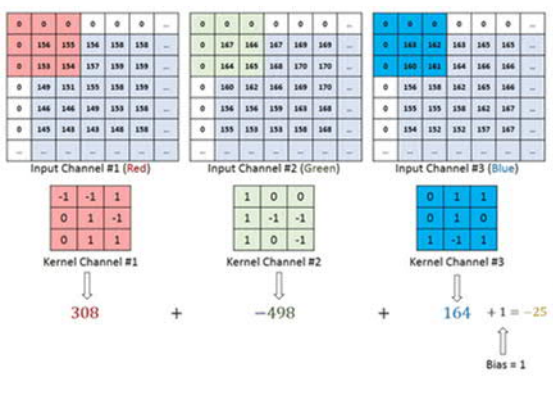




via https://haiicmu.github.io

373

373



CNN building blocks: Convolution in a color image

via https://haiicmu.github.io

374

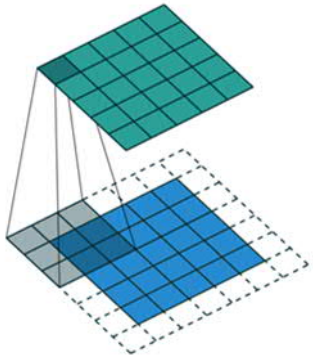
374

CNN building blocks: Padding

Each neuron "summarizes" its receptive field into a single value → convolutional layer will output *smaller* matrix

To keep an image from shrinking between layers, can use padding.

Add enough empty space to the input that it will "shrink" to be the actual image size.



via https://haiicmu.github.io

375

375

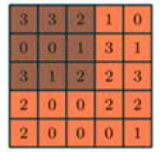
CNN building blocks: Pooling

Pooling is the opposite to padding.

A pooling layer takes a convoluted feature *after the convolution layer*, and shrinks it again.

Max pooling = just take the max value, throw out the rest.

Purpose: reduce computational complexity and prevent overfitting.



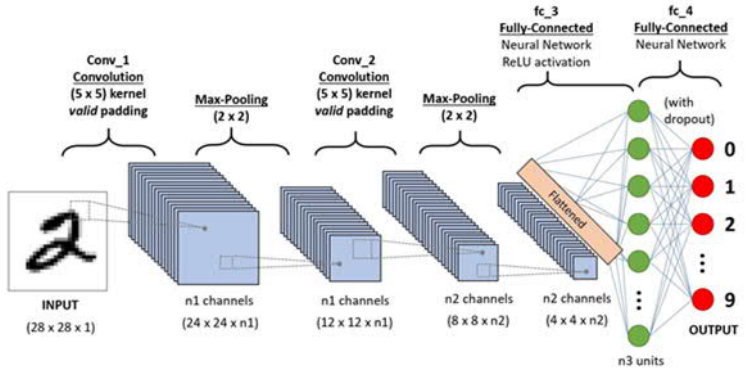
via https://haiicmu.github.io

via https://haiicmu.github.io

376

376

CNN: Architecture summary



via https://haiicmu.github.io

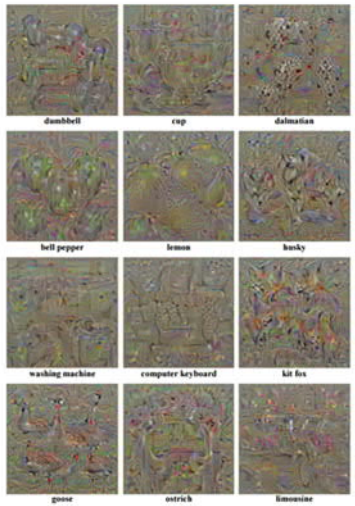
377

377

What are CNNs really

These gradient generated images are meant to be an image that will cause very high activation for the CNN to belong to that category.

More details:
https://distill.pub/2017/feature-visualization/



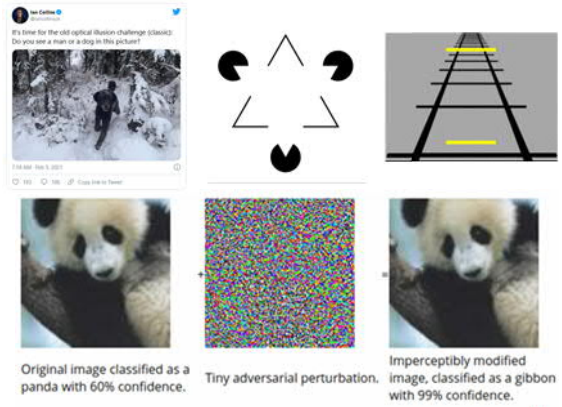
via https://haiicmu.github.io

378

How to trick a CNN vs a human

Trick a human

Trick a CNN



Nguyen, A., Yosinski, J., & Clune, J. (2015). Deep Neural Networks are Easily Fooled: High Confidence Predictions for Unrecognizable Images. *ArXiv:1412.1897 [Cs]*. http://arxiv.org/abs/1412.1897

Original image classified as a panda with 60% confidence.

Tiny adversarial perturbation.

Imperceptibly modified image, classified as a gibbon with 99% confidence.

379

Generative Adversarial Network: generating stuff

Fake paintings: Photo via Art and Artificial Intelligence Laboratory, Rutgers University

via https://haiicmu.github.io

380

380

This GAN-generated painting auctioned for $432,500 (2019)

'painting' = printout of a GAN output by French students

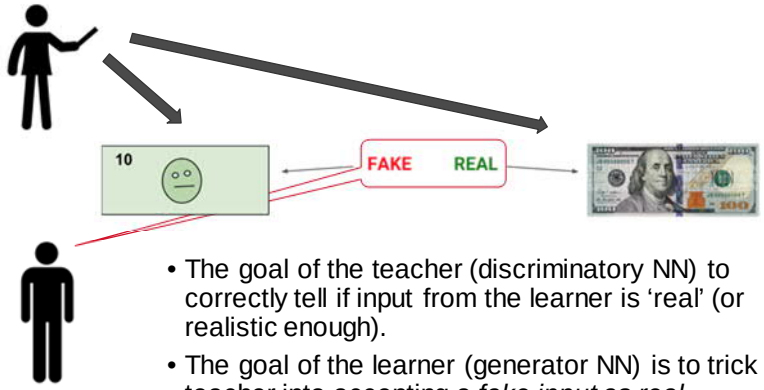
Code was borrowed: written as open-source by an american high schooler

A huge amount of excitement for generated art

via https://haiicmu.github.io

381

GAN: Generative Adversarial Network (2014)



- The goal of the teacher (discriminatory NN) to correctly tell if input from the learner is 'real' (or realistic enough).
- The goal of the learner (generator NN) is to trick the teacher into accepting a *fake input as real.*

via https://haiicmu.github.io

382

382

GAN: first train the teacher

Paintings by Peter Paul Rubens

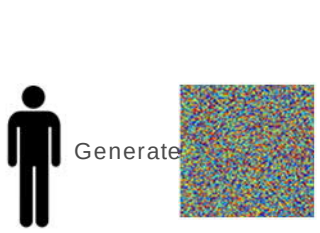
Not paintings by Peter Paul Rubens

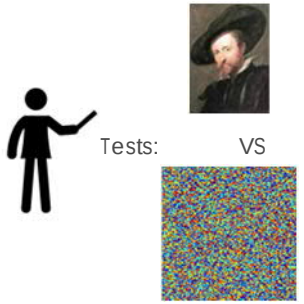
via https://haiicmu.github.io

383

383

GAN:
start with a learner who knows nothing





via https://haiicmu.github.io

384

384

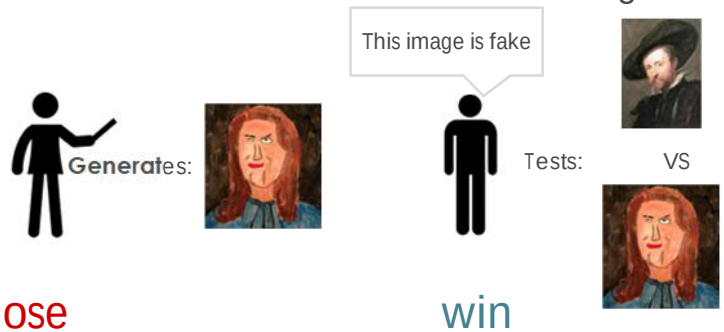
GAN:
start with a learner who knows nothing

via https://haiicmu.github.io

385

385

GAN:
start with a learner who knows nothing

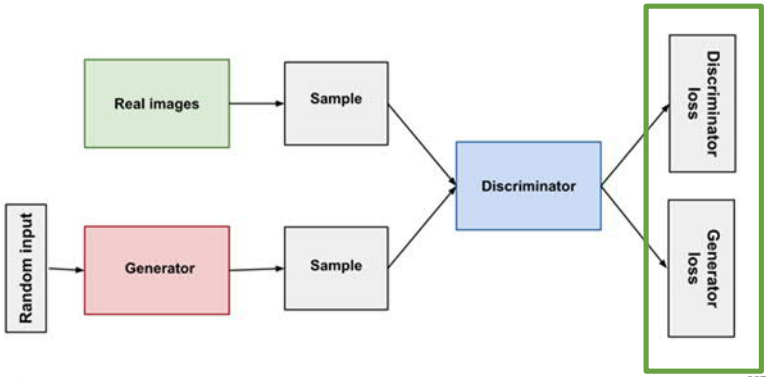


via https://haiicmu.github.io

386

386

Generative Adversarial Network Architecture



via https://haiicmu.github.io

387

387

GANs: How do the teacher and learner improve?

- The discriminator returns a *generator loss* which tells the learner how far from a true sample they were
- The discriminator also returns its own *discriminator loss* which helps it get better at detecting fakes

Each player gets a few attempts per round so that they each have a chance to learn.

***In practice, GANs are very difficult to get to stably converge

via https://haiicmu.github.io

388

388

Interactive demos: AI Experiments

- Fonts
  - A simple t-SNE for fonts http://fontmap.ideo.com/
  - Generative font hierarchies https://fontjoy.com/
- Cats
  - Old but good pix2pix https://affinelayer.com/pixsrv/
  - Autocomplete cats https://magic-sketchpad.glitch.me/
- Art generators
  - https://artbreeder.com/

For inspiration, take some time to play!

https://experiments.withgoogle.com/collection/ai

Zhu, J., Villareale, J., Javvaji, N., Risi, S., Löwe, M., Weigelt, R., & Harteveld, C. (2021). Player-AI Interaction: What Neural Network Games Reveal About AI as Play. *ArXiv:2101.06220 [Cs]*. http://arxiv.org/abs/2101.06220

via https://haiicmu.github.io

389

389

Cool applications of GANs

For further reading:

https://medium.com/@jonathan_hui/gan-some-cool-applications-of-gans-4c9ecca35900

https://machinelearningmastery.com/impressive-applications-of-generative-adversarial-networks/

Human-AI Co-Creation
(creative processes)

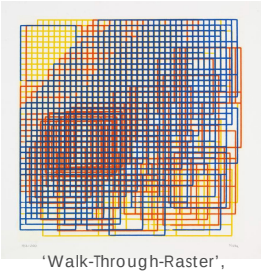
Example of Sketches to Color Photographs With pix2pix, Taken from Image-to-Image Translation with Conditional Adversarial Networks, 2016

390

AI & Art

- Again: various relationships relationship
  - AI to make art
  - Art making use of AI
  - Art reflecting on AI
  - AI reflecting on art?
- early algorithmic art, Nake et al ...

Virtual gallery: https://mlart.co/


'Walk-Through-Raster', by Frieder Nake, 1972

391

391

AI & Data Visualisation

https://youtu.be/I-ElVlHvHRM

392

392

AI & (Creative) Language

Tony Veale on Creative Machines:
https://www.youtube.com/watch?v=r-ToW5lestA

Veale, T., & Cardoso, F. A. (2019). *Computational Creativity: The Philosophy and Engineering of Autonomously Creative Systems*. Springer.

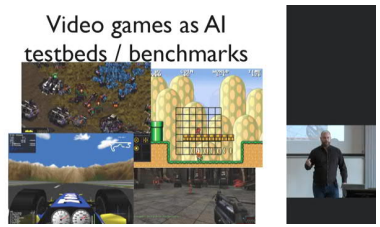
393

393

AI & Video Games

- Large space of applications in "generativity"
  - Note: often heuristics rather than "AI"

https://www.youtube.com/watch?v=U4m8mZOou6A



This Photo by Unknown Author is licensed under CC BY

Pfau, J., Smeddinck, J. D., Bikas, I., & Malaka, R. (2020). Bot or not? User Perceptions of Player Substitution with Deep Player Behavior Models. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–10. https://doi.org/10.1145/3313831.3376223

394

394

Art ~ Creativity (general problem-solving)

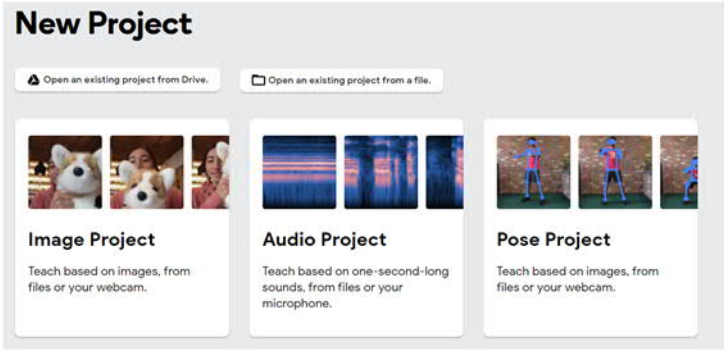
4 Experiments Where the AI Outsmarted Its Creators:
https://www.youtube.com/watch?v=GdTBqBnqhaQ

Building on that earlier work on AI outsmarting systems:
https://www.youtube.com/watch?v=Lu56xVlZ40M

395

395

Human-AI Co-Creation (development)



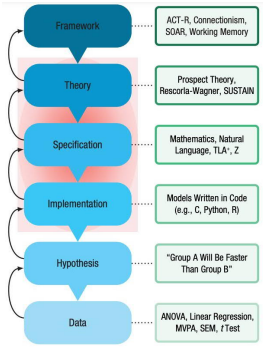
https://teachablemachine.withgoogle.com/train

396

396

Human-AI Co-Creation (research)

- Researching AI?
- A new kind of science?
- Luc Steels "evolutionary computational linguistics"



Guest, O., & Martin, A. E. (2021). How Computational Modeling Can Force Theory Building in Psychological Science. *Perspectives on Psychological Science*, 1745691620970585. https://doi.org/10.1177/1745691620970585

397

397

A Fighter Jet for the (Creative) Mind?

Meaningful Intelligence / Intelligent Augmentation: https://www.youtube.com/watch?v=C7D5EzkhT6A

398

398

Human-AI Interaction & Futures

3 Week Block Mode Lecture Series: Weeks 1 - 3 Summary

Jan Smeddinck

Unless otherwise noted, all materials: (CC BY-NC-SA 4.0)

399

399

Course Outline: Week 01

- Topic 01 - Introduction & Organisation
- Topic 02 - Basic Concepts of AI
- Topic 03 - History of AI
- Topic 04 - Human-Data Interaction
- Topic 05 - Data Visualisation & Communication
- Topic 06 - Human-AI Interaction Design

400

400

Course Outline: Week 02

- Topic 01 - Machine Learning Basics
- Topic 02 - Traditional Machine Learning Models
- Topic 03 - Machine/Deep Learning for Human Activity Recognition.
- Topic 04 - Research on Wearable-based Behaviour Analysis

401

401

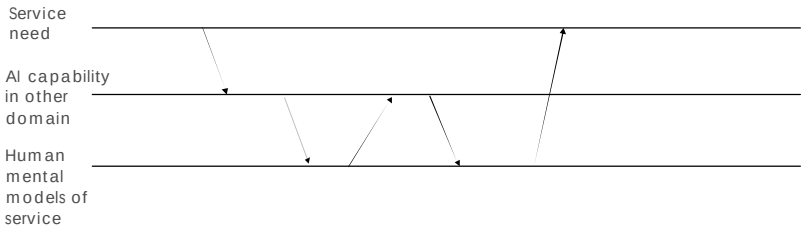
Course Outline: Week 03

- Topic 01 - Explainable, Interpretable & Relatable AI
- Topic 02 - AI Ethics
- Topic 03 - Humans-in-the-Loop
- Topic 04 - Recommender Systems
- Topic 05 - Conversational Interfaces
- Topic 06 - AI Agents & Robots
- Topic 07 - Human-AI Integration
- Topic 08 - Creative AI
- Topic 09 - Summary & Outlook

402

402

Ideating for AI



403

403

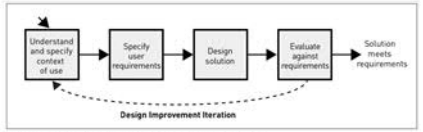


Figure 1. The UCD process as described in ISO 9241-210 [1].

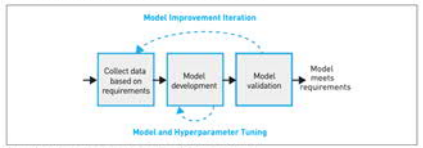


Figure 2. The machine learning train-validation-test cycle.

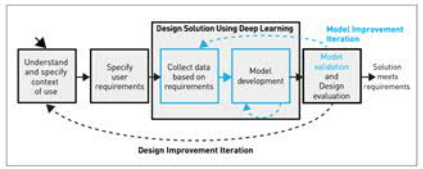


Figure 3. Adapted UCD process for deep learning in the context of interactive systems in HCI.

UCDDL

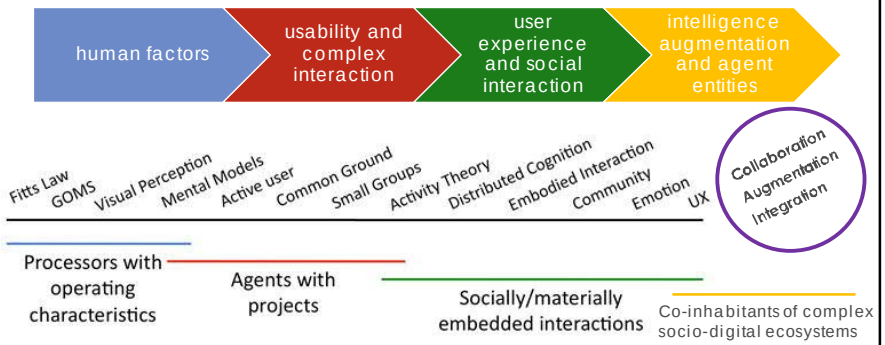
- User centred design process for deep learning
- Core idea:
  - Waterfall → iterative
  - And user-centred
- Work in progress:
  - Participatory and co-design with learning systems?

Le, H. V., Mayer, S., & Henze, N. (2021). Deep learning for human-computer interaction. *Interactions*, *28*(1), 78–82. https://doi.org/10.1145/3436958

404

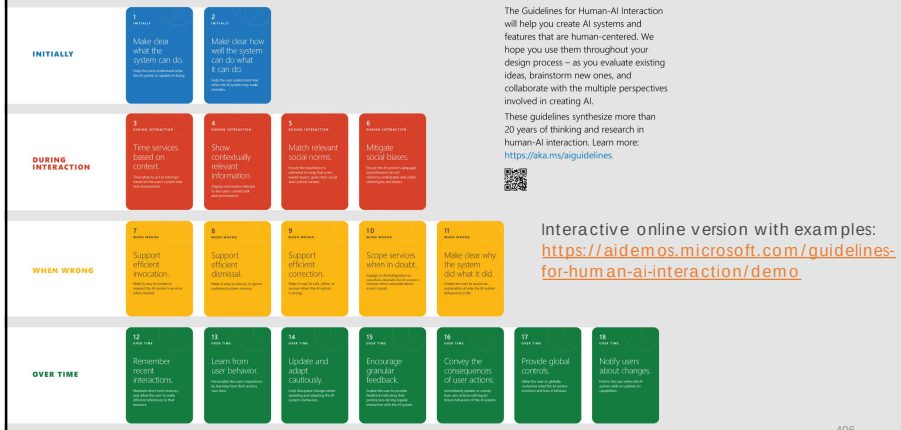
404

History of HCI + AI ???



via: https://www.interaction-design.org/literature/book/the-encyclopedia-of-human-computer-interaction-2nd-ed/human-computer-interaction-brief-intro

Guidelines for Human-AI Interaction



User Needs + Defining Success

People + AI Guidebook

- ① Find the intersection of user needs & AI strengths
  - Make sure you're solving a real problem
  - in a way where AI is adding unique value
  - Consider AI Principles and Responsible AI Practices
- ② Assess automation vs. augmentation
  - Automate tasks that are difficult or unpleasant
  - Augment bigger processes that people enjoy doing
- ③ Design & evaluate the reward function
  - how an AI defines successes and failures
  - deliberately design this function
  - optimize for long-term user benefits
  - limit potentially negative outcomes

Data Collection + Evaluation

- ① Translate user needs into data needs
  - What features, labels, and examples are needed?
  - Break down user needs, user actions, and ML predictions into necessary datasets
  - Formulate a plan to collect them
  - Inspecting data, identify potential bias sources, design data collection methods
- ② Source your data responsibly
  - Consider relevance, fairness, privacy, and security
  - Applies whether using existing dataset or building a new training dataset
- ③ Design for raters & labelling
  - Consider raters and the tools used for good label accuracy
- ④ Tune your model
  - Test and tune rigorously
  - Adjust parameters
  - Inspecting your data (often explains issues in outputs)

Mental Models

- ① Set expectations for adaptation
  - Identify and building on existing mental models
  - "What is the user trying to do?"
  - "What mental models might already be in place?"
  - "Does this product break intuitive patterns?"
- ② Onboard in stages
  - Set realistic expectations early
  - Describe user benefits, not technology
  - Describe the core value initially, introduce new features as they are used
  - Make it easy for users to experiment
- ③ Plan for co-learning
  - Connect feedback to personalization and adaptation
  - Fail gracefully to non-AI options when needed
- ④ Account for user expectations of human-like interaction
  - Communicate algorithmic nature and limits of system

409

409

Explainability + Trust

- ① Help users calibrate their trust
  - Instill user trust it in some situations, but indicate to double-check when needed
  - Articulate data sources
  - Tie explanations to user actions
- ② Optimize for understanding
  - Consider partial / full explanations, and/or process insights
- ③ Manage influence on user decisions
  - Consider communicating model confidence

410

410

Feedback + Control

- ① Align feedback with model improvement
  - Clarify differences between implicit and explicit feedback
  - Asking the right questions at the right level of detail
- ② Communicate value & time to impact
  - Understanding why people give feedback
  - Build on existing mental models to explain benefits
  - Communicate how/when user feedback will change experience
- ③ Balance control & automation
  - Help users control the aspects of the experience they want to
  - Easy opting out of giving feedback

411

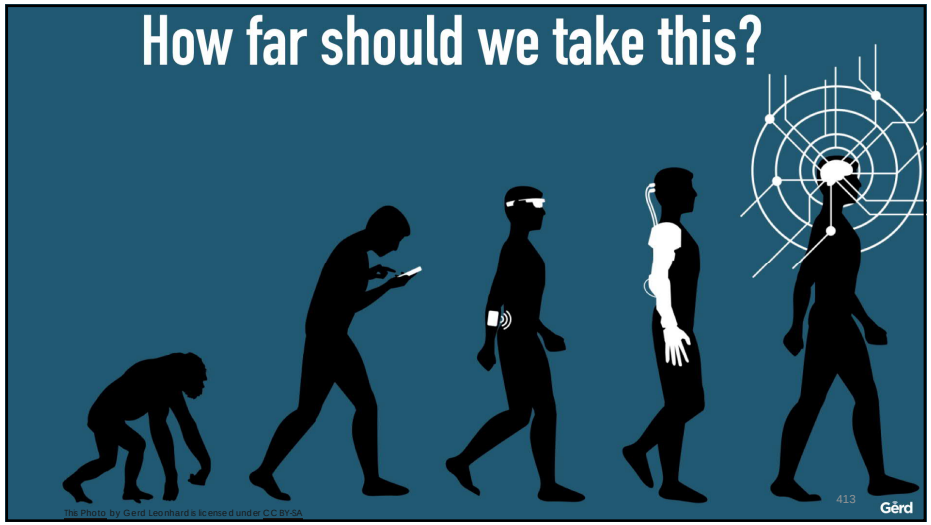
411

Errors + Graceful Failure

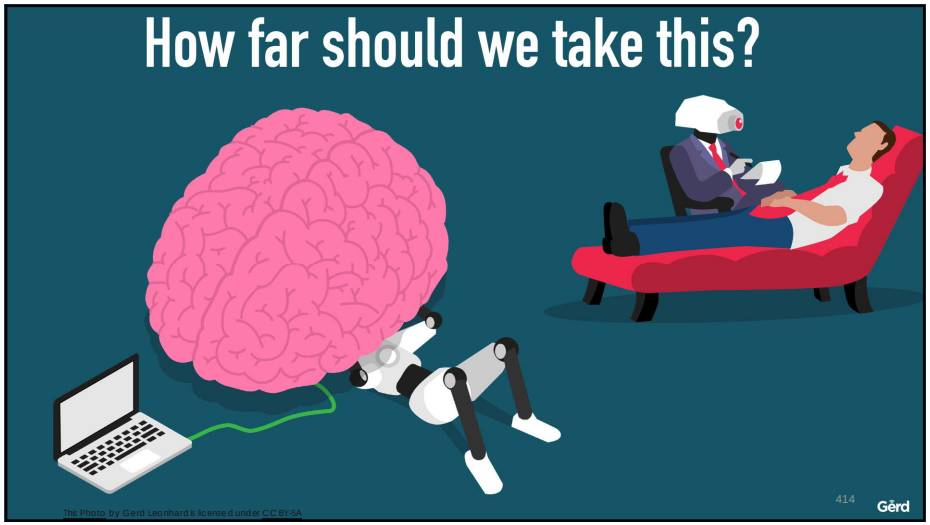
- ① Define "errors" & "failure"
  - System working as intended can still be perceived as failure
  - Acknowledge work-in-progress (if applicable)
- ② Identify error sources
  - Inherent complexity can make identifying source of errors challenging
  - Consider error / error source discovery strategies
- ③ Provide paths forward from failure
  - Probabilistic systems will fail at some point
  - Identify failure; user-centered discovery / resolve
  - Facilitate feedback and return to task

412

412

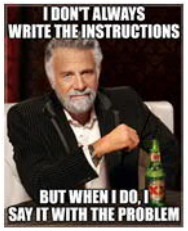


413



414

Human vs. AI Memes



Via https://imgflip.com/ai-meme?

415

An Apt Closing Statement

Luc Steels on Humane AI
https://www.youtube.com/watch?v=_MaQlwZlvs0

416

416